

UTILIZAÇÃO DE UM MODELO MATEMÁTICO-ESTATÍSTICO NO ESTUDO
DA RADIAÇÃO SOLAR E TEMPERATURA MÉDIA DO AR

ELZA CORREIA SUCHAROV

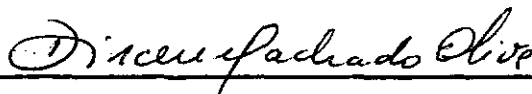
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE
PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JA
NEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO
GRAU DE MESTRE EM CIÊNCIAS (M.Sc.).

Aprovada por:



PEDRO GUERRERO SALAZAR

Presidente



DIRCEU MACHADO OLIVE



NILSON AUGUSTO VILLA NOVA

RIO DE JANEIRO, RJ - BRASIL
SETEMBRO DE 1978

Ao
Michael

Este momento me parece oportuno para ex
pressar a gratidão que tenho pela Sra. Nelly Balança que mui-
to me incentivou, concedendo-me, com sua compreensão e genere
sidade, condições e facilidades que me permitiram formar o a-
licerce de minha carreira universitária.

AGRADECIMENTOS

À COPPE, nas pessoas do seu Diretor Paulo Alcântara Gomes e do Chefe da Área de Recursos Hídricos, Rui Carlos Vieira da Silva.

À Universidade Federal do Rio de Janeiro pela oportunidade de desenvolver este trabalho.

Ao Prof. Pedro Guerrero pela sua eficiente orientação.

Ao Prof. Dirceu Machado pelas valiosas sugestões apresentadas no decorrer deste trabalho e juntamente com o Professor Nilson A. Villa Nova pelo exame crítico a que submeteram o manuscrito.

Ao Instituto Nacional de Meteorologia e Instituto Agrônomo de Campinas pelo fornecimento dos dados.

As amigas Cibele Gonçalves Picanço e Dulce Cardoso Washington pelo encorajamento nos momentos finais e, especialmente aos meus pais e meu irmão pelo carinho com que sempre acompanharam meus estudos.

RESUMO

Foi feito a revisão, sumário e discussão da análise da estrutura e modelos matemáticos usados para avaliação e remoção da periodicidade e dependencia nas séries de tempo geofísicas.

Para aplicação das técnicas descritas foram usados registros diários de radiação solar global e temperatura média diária do ar da Estação Centro Experimental de Campinas, S. Paulo (período 1965 a 1974) e da Estação Ecologia Agrícola, R. de Janeiro (período 1963 a 1973).

Na análise da periodicidade das séries foi empregada a Série de Fourier e na análise da dependencia foram aplicados modelos autoregressivos lineares até 6a. ordem. Periodicidades nos coeficientes de correlação $r_{k,\tau}$ da componente estocástica foram também analisadas. As componentes estocásticas independentes foram obtidas após terem sido removidas a periodicidade e dependencia das séries.

As funções de distribuição de probabilidades normal, lognormal 3-parâmetros, gama 3-parâmetros e gama "double branch" foram aplicadas na tentativa de ajustar a distribuição de frequências da componente estocástica independente.

ABSTRACT

The structural analysis and mathematical models used for evaluation and removal of the periodicity and dependence from the geophysical time series are reviewed, summarized and discussed.

Records of daily global solar radiation and mean air temperature in Campinas - São Paulo, Estação Centro Experimental de Campinas (period 1965 to 1974) and KM 47 - Rio de Janeiro, Estação Ecologia Agrícola (period 1963 to 1973) are used as the basic research data.

Fourier series are used to analyse the periodicity of the data. The linear autoregressive models are applied from the first to sixth order to describe the dependence of these series. Periodicities in serial correlation coefficients, $r_{k,\tau}$, of the stochastic components are also analyzed. Independent stochastic components are obtained by removing the periodicity and dependence from the series.

The probability distribution function normal, lognormal 3-parameters, gamma 3-parameters and double branch gamma are applied to fit the frequency distributions of the independent stochastic components.

ÍNDICE

I. INTRODUÇÃO

1.1 - Importancia da Meteorologia Estocástica	1
1.2 - Radiação Solar Global	3
1.3 - Objetivos do trabalho	5
1.4 - Organização do trabalho	5

II. REVISÃO DA LITERATURA

2.1 - Análise da estrutura das séries geofísicas ...	7
2.2 - Funções de distribuição de probabilidade para ajuste da distribuição de frequência da compo- nente estocástica independente	10

III. METODOLOGIA

3.1 - Conceitos e definições	13
3.1.1 - Processos estocásticos	13
3.1.2 - Estacionaridade	13
3.2 - Hipóteses básicas da análise da estrutura das séries temporais geofísicas	15
3.3 - Modelo Geral de uma série de tempo	16
3.3.1 - Tendencias e saltos	16
3.3.2 - Componente periódica	19
3.3.2.1 - Análise de Fourier	21
3.3.2.2 - Harmônicos Significantes	23
3.3.3 - Componente estocástica	26
3.3.3.1 - Função de autocorrelação	26
3.3.3.2 - Modelo de dependencia da componen- te estocástica	28
3.3.3.3 - Investigação da periodicidade nos coeficientes de autocorrelação e estimação dos coeficientes autore- gressivos do modelo	29

3.3.3.4 - Desvio padrão de $\zeta_{p,\tau}$	31
3.3.3.5 - Seleção da ordem do modelo autoregressivo	32
3.3.3.6 - Teste de adequacidade dos modelos autoregressivos	34
3.3.4 - Funções de distribuição de probabilidade teórica para ajuste da componente estocástica independente	35
3.3.4.1 - Função de distribuição de probabilidade normal	36
3.3.4.2 - Função de distribuição de probabilidade lognormal 3-parametros	36
3.3.4.3 - Função de distribuição de probabilidade gama 3-parametros	37
3.3.4.4 - Função de distribuição de probabilidade "double branch gamma"	38
3.3.4.5 - Teste para verificação de ajuste da distribuição de frequência	39

IV. APLICAÇÕES

4.1 - Séries empregadas e análise preliminar	41
4.1.1 - Introdução	41
4.1.2 - Avaliação dos dados disponíveis	41
4.1.3 - Preenchimento das falhas	43
4.2 - Emprego da metodologia	44
4.2.1 - Verificação da homogeneidade das séries	44
4.2.2 - Parametros estatísticos das amostras ..	44
4.2.3 - Componente periódica	46
4.2.4 - Análise da componente estocástica dependente e seleção do modelo autoregressivo	54
4.2.5 - Ajuste de uma distribuição de probabilidade à componente estocástica independente	54
4.2.6 - Discussão dos resultados	61

V. CONCLUSÕES E RECOMENDAÇÕES

I - INTRODUÇÃO

1.1 - Importancia da Meteorologia Estocástica

A maioria dos processos de tempo que ocorrem na natureza, observados em intervalos inferior a um ano, podem ser considerados processos periódicos-estocásticos, podendo a periodicidade ser atribuída aos ciclos astronômicos e a estocasticidade às várias fontes de aleatoriedade dentro do sistema Terra-Atmosfera. A atmosfera é a maior fonte de estocasticidade nos processos geofísicos correlatados com a incidência de radiação solar, embora os oceanos, a superfície da terra, e a crosta terrestre também produzam estocasticidade.

Algumas das fontes primárias de estocasticidade em processos geofísicos são: a turbulência e vorticidade em grande escala associada com o movimento de fluidos; a aleatoriedade na transferência de calor, ou geralmente, na transferência de energia; os efeitos aleatórios nos processos de transporte de material em suspensão, flutuantes, ou de outra forma; a aleatoriedade associada a passagem da água através os pontos críticos de congelamento, evaporação, condensação e sublimação; e os efeitos de fatores aleatórios em estados permanente do meio, incluindo mudanças biológicas.

Assim sendo, observa-se que sem a meteorologia estocástica, torna-se difícil um bom entendimento dos processos estocásticos geofísicos e surge então a necessidade premente de se desenvolverem pesquisas na área de meteorologia estocástica, principalmente no Brasil, onde quase nada existe a respeito.

Face a essa necessidade e tendo em vista a importância do conhecimento do comportamento da parte aleatória da radiação solar nos processos geofísicos, além da ênfase dada nos últimos anos ao aproveitamento da radiação solar devido a crise mundial do petróleo e os crescentes problemas ecológicos proveniente da poluição produzida pelas fontes convencionais de energia, é de grande valia contribuir-se tanto para o desenvolvimento da meteorologia estocástica como para os estudos sobre radiação solar, que vem a ser a realização de um processo periódico-estocástico.

Por outro lado, para possível aproveitamento da radiação solar, torna-se necessário o conhecimento de sua distribuição espacial. Em presença da deficiência de registros deste parametro, tem sido desenvolvidos vários estudos práticos-teóricos objetivando a estimação da radiação solar a partir de outros parametros meteorológicos, quase sempre através de retas de regressão linear. Nesse sentido, o estudo de outros parametros meteorológicos como base para estimação da radiação solar incidente, a luz dos processos estocásticos, constitui uma grande contribuição para o enriquecimento da meteorologia estocástica e fortalece os esforços que o País vem empenhando para acele

rar os estudos e pesquisas sobre radiação solar. Objetivando seguir esta linha, escolheu-se a temperatura média do ar como a variável a ser relacionada com a radiação solar. Esta escolha foi baseada no fato de que a radiação solar tem influencia significativa no regime da temperatura média do ar e na disponibilidade e confiabilidade dos dados de temperatura. Entretanto, se tem em mente, que melhores resultados poderiam ser encontrados para estimação da radiação solar com o emprego de horas de brilho solar ou mesmo de temperaturas máximas.

1.2 - Radiação Solar Global

A radiação solar que incide em um plano horizontal situado no limite superior da atmosfera é calculada por:

$$I = I_0 \cos Z$$

onde I_0 é a constante solar (igual a $1.94 \text{ cal/cm}^2 \cdot \text{min}$ conforme Paltridge e Platt, 1976) e Z_0 é a distancia zenital do sol.

Ao atravessar a atmosfera, os raios solares podem ser refletidos, absorvidos e dispersados pelas nuvens, gases e aerossóis. A parte da radiação solar que é dispersada e alcança a superfície da Terra é conhecida como radiação difusa. A radiação difusa mais a radiação direta (radiação solar que não sofre alterações ao atravessar a atmosfera) constituem a radiação solar global que é o total de energia incidente na superfície da Terra em onda curta.

A medida da radiação solar global é feita por piranômetros. Estes são instrumentos basicamente constituídos de três partes compreendendo um corpo maciço, uma cúpula de quartzo e um sensor. Os dois tipos de sensores empregados são os constituídos de termopilhas e o fotovoltaico, sendo este último de emprego mais recente.

No primeiro caso o sensor é constituído de um número determinado de termopilhas (de acôrdo com o fabricante) e, geralmente, de constant-manganina. O sensor fotovoltaico é uma plaqueta de material semi-condutor, silício por exemplo, revestida com uma camada de cobre na parte superior e inferior e são conhecidas como células solares. Em ambos os casos a radição solar global é função da diferença de potencial nos sensores.

As unidades usuais empregadas para medida da radiação solar são a caloria por centímetro quadrado por minuto, ($\text{cal/cm}^2 \text{ min}$) e o Watt por metro quadrado (W/m^2).

1.3 - Objetivos do trabalho

O objetivo principal deste trabalho é contribuir para o desenvolvimento da meteorologia estocástica no Brasil. Para atingí-lo, propõe-se a:

- (1) Obter um modelo autoregressivo linear para as componentes estocásticas das séries diárias de radiação solar global

e temperatura média do ar, empregando a metodologia proposta por Yevjevich, 1972a.

- (2) Condensar as informações contidas na componente estocástica independente obtida com o emprêgo de modelos autoregressivos pelo ajuste apropriado de uma função de distribuição de probabilidade à sua distribuição de frequência.
- (3) Obter um modelo para estimação da radiação solar global em função da temperatura média do ar, empregando os modelos autoregressivos lineares encontrados.

1.4 - Organização do trabalho

O Capítulo II deste trabalho apresenta, de forma sucinta, a revisão da literatura apropriada. No Capítulo III está descrita uma metodologia de trabalho para análise da estrutura das séries meteorológicas de tempo e as funções de distribuição de probabilidade mais empregadas para o ajuste da componente estocástica independente. O Capítulo IV versa sobre a aplicação da metodologia apresentada no Capítulo II às séries de radiação solar global e temperatura média diárias para as Estações meteorológicas Centro Experimental de Campinas localizada em Campinas, Estado de São Paulo e Ecologia Agrícola localizada no km 47 da Rodovia Presidente Dutra, Estado do Rio de Janeiro. O Capítulo V, e último, apresenta as conclusões e recomendações do trabalho.

II - REVISÃO DA LITERATURA

2.1 - Análise da estrutura das séries geofísicas

Grenander e Rosenblat (1957), Hannam (1970), Anderson (1971) e Yevjevich (1972b) entre outros, são os principais estudiosos da teoria dos processos estocásticos, tendo este último se dedicado à investigação de processos estocásticos hidrológicos e meteorológicos.

Em Yevjevich (1972a) foi apresentada uma metodologia para análise da estrutura das séries temporais geofísicas que são realizações de processos estocásticos geofísicos. Após terem sido removidas as tendências e os saltos que se identificarem, estes processos podem ser considerados estacionários no caso de séries anuais. Entretanto, no caso de séries mensais, semanais ou diárias, não se pode desprezar a periodicidade presente em alguns.

A investigação da periodicidade em séries temporais foi iniciada por Schuster (1898) através do uso de periodogramas. Mais tarde, Tuckey (1950) e Bartlett (1950) deram grande impulso à análise espectral. A transformada rápida de Fourier (FFT) na análise do espectro de séries temporais permitiu estimar de forma mais rápida o espectro. Machado (1970) empregou a FFT na análise de espectro de séries de vazões diá-

rias, o que abriu novas perspectivas na análise da estrutura de séries geofísicas no Brasil.

Rodriguez-Iturbe (1971) também contribui para a investigação da periodicidade nas séries temporais utilizando um conjunto de espectros variáveis ao longo do tempo. Nessa abordagem ele emprega as técnicas de modulação sugeridas por Tuckey (1961). Yevjevich (1972a) usou a análise harmônica para as estatísticas das séries e apresentou um procedimento para o teste de harmônicos significantes usando a soma da variancia explicada por todos os harmônicos e dois valores críticos pré-definidos. Sales (1977) desenvolveu um novo teste para a determinação do número de harmônicos significantes em parâmetros periódicos baseado na teoria dos intervalos de confiança.

A metodologia apresentada por Yevjevich (1972a) para análise da estrutura das séries observadas dos processos estocásticos geofísicos estacionários pode ser sumariada em quatro etapas principais:

1. determinação da componente periódica, após ter sido pesquisada a presença de tendências e saltos na série, e eliminação dos mesmos caso existam;
2. remoção da periodicidade nas estatísticas da série e obtenção da componente estocástica dependente ou independente;
3. ajuste do modelo de dependencia autoregressivo à componente estocástica dependente e obtenção da componente independente;

4. ajuste de uma distribuição de probabilidade à componente in dependente;

Essa metodologia vem sendo bastante empregada por vários pesquisadores, e dos resultados encontrados pode ser in dicado que:

A - Séries de precipitação e vazão:

As séries de precipitação anual, precipitação anual efetiva (precipitação menos evaporação) e vazão anual são aproximadamente séries de tempo estacionárias. As séries de precipitação anual são aproximadamente séries independentes e as séries anuais de precipitação efetiva e vazão podem ser tan to dependente como independente. Quando são dependentes, a de pendencia pode ser aproximada por um modelo linear autoregressivo de 1a. ou 2a. ordem (Yevjevich, 1963 e 1964).

As séries mensais de precipitação, precipitação e fetiva e vazão tem periodicidade de doze meses na média e desvio padrão. Após essas periodicidades terem sido removidas, a parte restante pode ser considerada aproximadamente como um pro cesso estocástico estacionário independente para precipitação mensal e no caso de vazões mensais, um processo estocástico es tacionário linearmente dependente (Roesner e Yevjevich, 1966). No Brasil, as pesquisas com séries de precipitação e vazão men sais tem sido realizadas por Lattari e Guerrero (1975), Eid (1976), Sales (1977), e outros.

As séries diárias se enquadram entre os mais complexos processos geofísicos de tempo, embora modelos similares aos anteriores possam ser obtidos. O problema nos estudos com séries diárias está na dificuldade de ajustamento de uma função de distribuição de probabilidade à componente aleatória. Pesquisas com séries de vazão diárias feitas por Quimpo (1967) mostraram que a variancia explicada obtida pela aplicação dos modelos autoregressivos de 1a. a 3a. ordem não diferem apreciavelmente da variancia explicada quando do uso do modelo de 3a. ordem. Machado (1970) usou o modelo autoregressivo de 6a. ordem para séries de vazões diárias e Kottegoda (1972) um modelo de 4a. ordem para descrever a dependencia de vazões para um período de 5 dias. Kelman (1976) desenvolveu um modelo geral para descrição e geração de séries de precipitações e vazões diárias que apresentou resultado satisfatório quando testado pelo emprego de registros de várias localidades dos Estados Unidos.

B - Séries de Radiação Solar

Só foi possível encontrar uma referencia na literatura disponível sobre análise da estrutura de séries de radiação solar diária (Yevjevich, 1971). Neste trabalho foi analisada a estrutura das séries de radiação solar diária de 14 estações situadas no Oeste dos Estados Unidos, objetivando demonstrar o grau e tipos de estocasticidade criada pela atmosfera neste processo. Os valores encontrados para os coeficientes de autocorrelação com defasamento de um dia da componente estocástica dependente para as 14 estações foram da mesma ordem de grandeza que os encontrados para séries de precipitação diária,

e com base nos correlogramas obtidos, Yevjevich assumiu que o modelo autoregressivo linear de 1a. ordem era suficiente, em primeira aproximação, para representar o modelo de dependência físico. A distribuição empírica da densidade de frequência da componente independente para as 14 séries mostraram-se negativamente assimétricas.

C - Séries de Temperatura do Ar

Utilizando a análise harmônica e espectral Panosfsky (1955), Griffth et al (1956), Julian (1967), Kothandaraman (1971), apontaram o ciclo anual em séries de temperatura do ar como o único verdadeiramente significativo. Ainda nesta linha, Velz e Gannon (1960) colocaram em evidencia o ciclo anual de temperatura do ar e de outros elementos meteorológicos afim de chegar ao ciclo da temperatura da água. Cluis (1972) utilizou a temperatura residual do ar após ter sido definida a característica climática senoidal em um modelo que gera a temperatura residual da água.

Song e outros (1973) em seu trabalho pesquisou a persistencia sobre dois dias consecutivos da componente estocástica residual de séries de temperatura diária do ar e a modelou por um processo autoregressivo de 1a. ordem. Llamas e Triboulet (1976), analisando a estrutura de séries de temperatura diária do ar empregou 3 harmônicas para determinar o comportamento periódico e utilizou o modelo autoregressivo linear de 2a. ordem para descrever a componente estocástica. Furman (1978) trabalhando com séries de temperatura máxima diária tam

bem empregou o modelo autoregressivo linear de 2a. ordem na descrição da dependência da componente estocástica.

2.2 - Funções de distribuição de probabilidade para ajuste da distribuição de frequência da componente estocástica independente

A função densidade de probabilidade normal tem sido usada por muitos pesquisadores como a distribuição de probabilidade teórica da componente aleatória de processos estocásticos. No campo dos processos estocásticos geofísicos é frequente o uso de funções de densidade de probabilidade lognormal e gama. Markovic (1965) verificou que as funções densidade de probabilidade normal, lognormal e gama oferecem bom ajuste para a componente estocástica independente de séries anuais de precipitação e vazão. Eid (1976) conseguiu bom ajuste empregando a função densidade de probabilidade lognormal (3 parâmetros) com séries de precipitação mensal. Bonne (1971) empregou as funções de distribuição lognormal, gama e Pearson Tipo III logaritmica para o ajuste da distribuição de frequência de vazões mensais. Tao (1973) trabalhou com as séries de vazões diárias utilizadas anteriormente por Quimpo (1967) e concluiu que as "caudas" das distribuições da componente independente são geralmente longas podendo serem aproximadas por funções exponenciais simples. Das várias funções de distribuição de probabilidade nenhuma passou no teste de qui-quadrado sendo porém a distribuição gama "double-branch" a que apresenta menor valor para o mesmo.

Em outro trabalho posterior, Tao e outros (1976) usaram séries de vazões diárias e séries formadas pelas médias de 3 dias, 7 dias, 13 dias e um mês. Foram testados os ajustes das distribuições das componentes estocásticas independentes η , ξ , ζ obtidas respectivamente com o emprego das vazões históricas, logarítimos das vazões históricas e logarítimos da componente estocástica dependente. Os resultados podem ser sumarizados em: - as funções gama e lognormal são as que melhor se ajustam a séries mensais, sendo que a função normal modificada por 3 ou 4 termos dos polinômios de Hermite também dá um bom ajuste; - para as séries de 13 dias, a distribuição de frequencia de η e ζ são bem ajustadas pelas funções de densidade de probabilidade lognormal, gama e normal modificada por 3 ou 4 termos do polinômio de Hermite. Para as séries ξ as funções que melhores se ajustaram foram a gama "double-branch" e a combinção de Pearson Tipo VII e gama; - para as séries de 7 dias a distribuição de frequência das séries ξ e ζ somente se ajustam à função gama "double-branch" e à combinação da função Pearson tipo VII e gama, sendo que no último caso apenas 9 das 17 séries empregadas deram bom resultado; no caso da distribuição de frequência da série η , a função lognormal, normal modificada por 3 ou 4 termos do polinômio de Hermite e a gama modificada pelos 3 termos do polinômio de Laguerre ajustam-se em um terço dos 17 casos, enquanto a função gama "double-branch" ajustam bem 10 séries, e a combinação das funções normais e gama aplicaram-se a 8 casos; - para as séries de 3 dias, a função gama "double-branch" ajustou bem em 8 das séries de η ; entretanto, nenhuma das funções estudadas ajustaram-se bem à distribuição de frequência da ξ e η ; - para séries diárias, tam

bém não é encontrado nenhum ajuste com as distribuições estudadas, sendo que a função gama "double-branch" é a que apresenta o menor valor no teste de qui-quadrado.

Llamas e outros (1976) considerou a componente estocástica independente de séries de temperatura média diária normalmente distribuída para geração de uma nova série com as mesmas características estatísticas e sequenciais e Furman (1978) usou a função de distribuição Beta-P-3 parâmetros para a juste da componente independente de série de temperaturas máximas diárias que passam no teste de qui-quadrado para altos níveis de probabilidade.

Dos trabalhos realizados neste sentido pode ser destacada a grande dificuldade de ajuste da distribuição da componente estocástica independente quando o intervalo de tempo entre as observações vai diminuindo. Está demonstrado que nenhum problema de ajuste existe com séries anuais, poucos com séries mensais e a partir daí um aumento de dificuldade considerável, chegando a uma maior dificuldade no caso de séries diárias.

III - METODOLOGIA

3.1 - Conceitos e Definições

3.1.1 - Processos Estocásticos

Designemos por S um sistema físico. O estudo da evolução deste sistema, no tempo, pressupõe dois princípios fundamentais, Costa, 1971:

1º Princípio: Dado um sistema físico S devemos sempre considerar as influências exteriores sobre S . Quaisquer que sejam os efeitos gerados, em intervalos de tempos considerados, por influência de corpos exteriores, estes efeitos se acumulam e constituem as causas das variações de configurações de S .

2º Princípio: Dadas duas configurações assumidas por S , existe sempre uma probabilidade associada à mudança da primeira configuração à segunda em dado tempo. Estas probabilidades constituem base para cálculos.

Estas noções introdutórias permitem apresentar a seguinte conclusão:

Chama-se processo estocástico ou aleatório, aquele para o qual, a intervalos aleatórios ou determinados de tempo, correspondem mudanças de estado de S associadas à leis de probabilidade.

Uma série discreta de tamanho N ou uma série continua de comprimento T , é considerada a realização de um proceso estocástico.

3.1.2 - Estacionaridade

Para definir estacionaridade é usado um conjunto de p séries, tal como mostrado na figura (III.1), Yevjevich, (1972b).

O processo $X(t)$ onde t representa o tempo, é essencialmente caracterizado pela família de funções de distribuição $F\{x(t_1), F x(t_2), \dots, F x(t_p)\}$ das variáveis aleatórias $X(t_1), X(t_2), \dots, X(t_p)$.

O conceito de estacionaridade é então o seguinte: Se as funções de distribuição são idênticas para qualquer t , ou seja, são independentes de t , o processo é estritamente estacionário. Em caso contrário é evolutivo.

Como a realização de um processo que se enquadre totalmente na definição acima é praticamente impossível, recorrer-se a uma definição menos rígida de estacionaridade. Essa definição é limitada ao estudo das propriedades determinadas pe-

AMOSTRAS

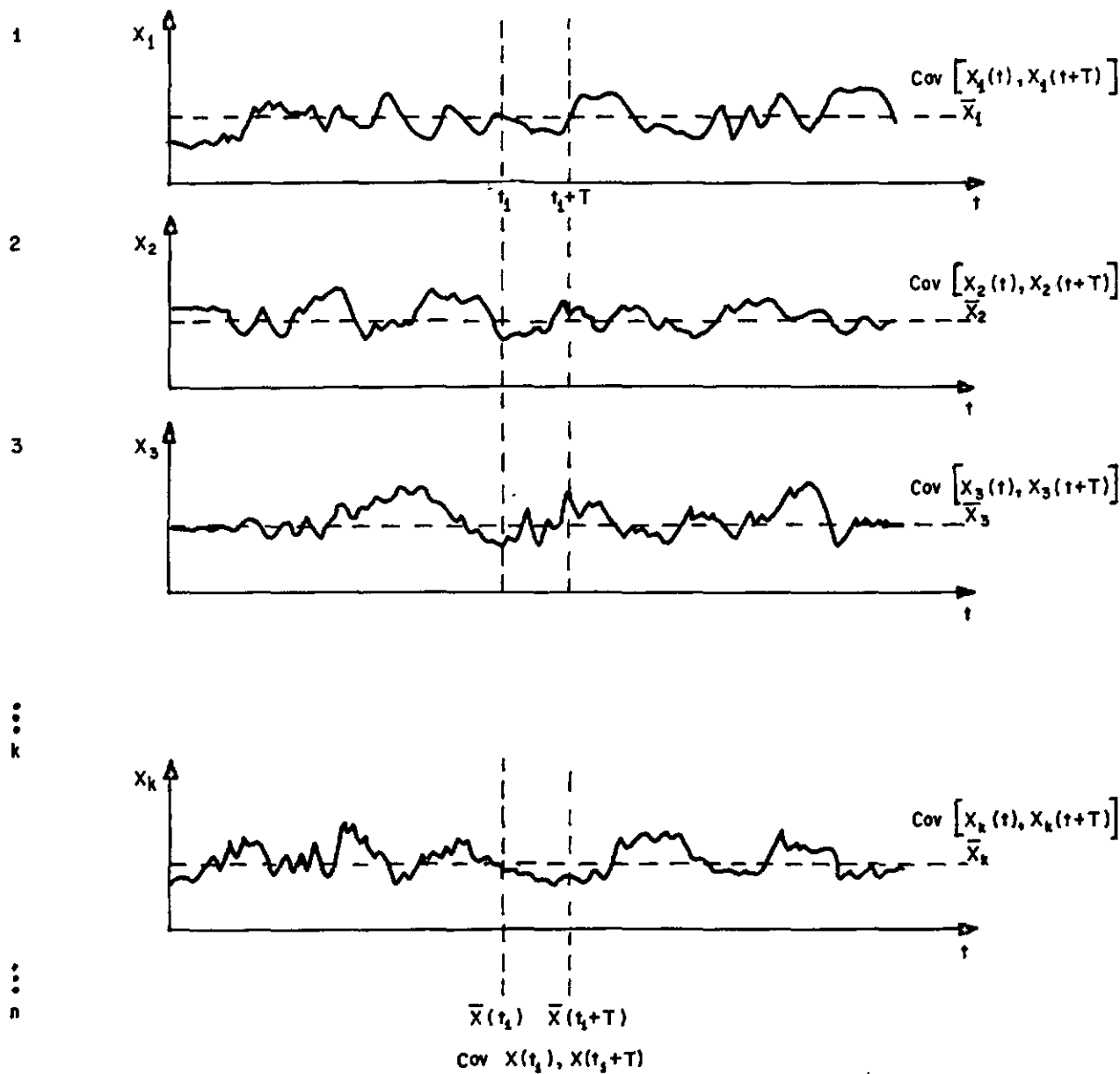


FIGURA III.1 - SERIES CRONOLÓGICAS

los momentos da função de distribuição e não pelas funções de distribuição propriamente ditas.

A estacionaridade implica em que os momentos da função de distribuição sejam independentes de t . De uma maneira geral, se $\mu(p_i)$ é o p -iésimo momento da população a qual $X(t_i)$ pertence e $m(p_i)$ o momento da amostra observada de tamanho N , o processo é estacionário de ordem p se ocorrer convergência na probabilidade de $m(p_i)$ para $\mu(p_i)$ qualquer que seja i , quando n tende a infinito:

$$P\{m(p_i) - \mu(p_i) \leq \varepsilon\} = 1 \quad \forall \varepsilon > 0 \quad \text{quando } N \rightarrow \infty$$

Na prática, no caso de séries meteorológicas, a análise da estacionaridade é limitada à ordem dois. Se diz então que o processo é estacionário de segunda ordem ou fracamente estacionário, ou ainda estacionário no sentido amplo, se os dois primeiros momentos, média e covariância, são independentes do tempo. Qualquer que seja t , a média e covariância são tais que:

$$E\{X(t)\} = \mu = \text{constante}$$

$$\text{cov}\{X(t), X(t+\tau)\} = E\{(X(t)-\mu)(X(t+\tau)-\mu)\} = f(\tau)$$

com τ sendo o defasamento.

3.2 - Hipóteses básicas da análise da estrutura das séries temporais geofísicas

O desenvolvimento de modelos matemáticos para a descrição das séries de tempo geofísicas e a estimação dos parâmetros do modelo representam uma forma avançada de extração, condensação e descrição das informações contidas nos dados observados. Entretanto, para aperfeiçoamento desses modelos são necessários métodos e esses métodos não podem ser desenvolvidos sem hipóteses postuladas. Essas hipóteses são frequentemente desenvolvidas ao longo da experiência com um grande número de séries, das propriedades físicas do processo e do entendimento geral do fenômeno. As hipóteses fundamentais que acompanham a análise da estrutura das séries temporais geofísicas são:

- (1) as séries temporais geofísicas podem ser separadas em componente periódica e componente estocástica sem prejuízo para o entendimento final e descrição da estrutura da série de tempo e extrações de informação;
- (2) toda variação aleatória da série está praticamente contida na componente estocástica, enquanto os erros de amostragem inevitáveis são deixados dentro dos parâmetros estimados dos parâmetros periódicos;
- (3) a componente estocástica da série é aproximadamente uma variável aleatória estacionária de uma certa ordem de estacionaridade;

- (4) a não homogeneidade e inconsistência nas séries são detectadas, descritas e removidas, se necessário, a priori, na análise da estrutura das séries;
- (5) a análise da estrutura das séries e sua descrição matemática não perpetuam erros amostrais;
- (6) a análise de um conjunto de séries em uma região pode melhorar significativamente os modelos e coeficientes e/ou parâmetros estimados;
- (7) a seleção do intervalo de tempo (1 dia, 3 dias, 1 mês, etc) não afeta a aplicabilidade dos métodos desenvolvidos; e
- (8) o número de parâmetros estimados a partir da amostra deve ser mínimo, afim de que o mínimo de graus de liberdade restante não seja muito diminuído.

3.3 - Modelo Geral de uma série de tempo

Aproximadamente toda série meteorológica constituída por valores mensais ou por intervalos de tempo menores é uma série composta de elementos determinísticos (periódicos-senoidais) e estocásticos (estacionários-ergódicos) com superposição, algumas vezes, de componentes transientes.

Se $X(t)$ é uma série meteorológica tomada em intervalos de tempo iguais, o modelo heurístico geral para descrever esta sequência pode ser dado por:

$$X_t = T_t + P_t + \epsilon_t \quad (\text{III.1})$$

onde,

- T_t é a componente transiente,
- P_t é a componente periódica ou cíclica, e
- ϵ_t é a componente estocástica ou aleatória

Caso não hajam tendências e/ou saltos a equação (III.1) pode ser escrita como a soma de uma componente periódica e uma componente estocástica:

$$X_t = P_t + \epsilon_t \quad (\text{III.2})$$

3.3.1 - Tendências e saltos

A tendência é definida como uma mudança sistemática e contínua sobre uma série longa em qualquer parâmetro da série. A inconsistência (erros sistemáticos) e a não homogeneidade, saltos (mudanças na natureza, bruscas, causadas tanto pelo homem como por processos naturais) são os fatores principais no aparecimento de tendências a longo prazo.

Se a partir dos valores da série a ser analisada fazemos um gráfico, podemos identificar qualitativamente se existe tendência e qual o tipo de tendência apresentada ou seja: se crescente, decrescente ou também identificar qualitativamente a possível existência de saltos. Para fortalecer as conclusões do exame gráfico quanto a presença de uma tendência, pode-se empregar o teste de

Spearman (Mitchell et al., 1966). O teste consiste em definir uma estatística para testar a hipótese do caráter aleatório de uma série contra a hipótese de tendência.

Seja uma série de N valores tendo cada valor uma ordem sequencial i e um número de ordem crescente k_i . Para $N > 8$ a estatística

$$st = r_s \sqrt{\frac{N-2}{1-r_s^2}} \quad (\text{III.3})$$

segue uma distribuição de Student com $(N-2)$ graus de liberdade. A estatística r_s é definida por:

$$r_s = 1 - \frac{\sum_{i=1}^N \sigma_i^2}{N(N^2-1)} \quad (\text{III.4})$$

onde $\sigma_i = k_i - 1$

O valor de st calculado é comparado com o valor teórico correspondente a um nível de significância α . Se $st \leq st(1-\frac{\alpha}{2})$ a hipótese da série aleatória é aceita e então não há presença de tendências na série.

Outro teste bastante utilizado para detectar tendências ou saltos é o teste de constância de parâmetros. Nesse teste a série é dividida normalmente em duas sub-séries e a seguir são aplicados os testes correspondentes. Para o caso das médias, o teste a ser empregado é o teste de Student. O teste de Student consiste em determinar a

estatística t da seguinte forma, (Neville e Kennedy, 1964):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left\{ \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 \right\} \cdot \frac{n_1 + n_2}{n_1 n_2}}} \quad (\text{III.5})$$

onde,

n_1 e n_2 são os tamanhos das subamostras,

\bar{X}_1 e \bar{X}_2 são as médias das subamostras obtidas através do 1º momento em torno da origem,

X_{1i} e X_{2i} são os valores que compõem as subamostras

A estatística t segue a distribuição de Student com $N-2$ graus de liberdade, sendo $N=n_1+n_2$. Para que a hipótese de constância da média seja aceita, a estatística t calculada deve ser menor do que um valor crítico t_c obtido das tabelas de distribuição de Student, para um determinado nível de significância α . Usualmente, o nível de significância α especificado para rejeitar a hipótese de constância das médias é 5% e 1%. Devido a hipótese básica do teste de Student ou seja, que as duas amostras pertençam a mesma população, as duas variancias estimadas devem ser consistente com esta hipótese, isto é, as duas variancias não devem ser significativamente diferentes. Isto pode ser verificado por meio do teste de Fisher, que consiste na definição da estatística F dada por:

$$F = \frac{Z_1}{Z_2}, \quad \text{para } Z_1 \geq Z_2 \quad (\text{III.6})$$

onde Z_1 é a variancia da amostra de tamanho n_1 e Z_2 a variancia da amostra de tamanho n_2 . A estatística F segue a distribuição de Fisher com $\alpha_1 = n_1 - 1$ e $\alpha_2 = n_2 - 1$ graus de liberdade. Da mesma forma que o teste de Student, se o valor de F calculado pela equação (III.6) não exceder o valor crítico F_c obtido das tabelas de distribuição de Fisher para um determinado nível de significancia (5% ou 1%) a hipótese de constancia da variancia é aceita.

Comprovada a tendencia, a mesma pode ser avaliada por uma série de N valores, com a ajuda de um polinômio de grau p, da forma:

$$T(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \dots + \lambda_p t^p \quad (\text{III.7})$$

onde, $1 \leq p \leq N-1$.

Quando p é relativamente pequeno em relação a (N-1), os coeficientes $\lambda_0, \dots, \lambda_p$ podem ser determinados pelo método dos mínimos quadrados. Frequentemente, as tendencias encontradas são lineares e então estimadas por polinômios do 1º grau. Caso contrário, a aproximação por uma parábola ou uma curva do 3º grau dá bons resultados, (Chakravarti e outros, 1967). Se a tendencia apresenta-se mais complexa, o uso de polinômios de grau elevados nem sempre é justificado, e então a avaliação da tendencia pelo emprego da técnica das médias móveis torna-se mais adequada.

3.3.2 - Componente periódica

A investigação da existencia de periodicidade nos parametros de uma série de tempo pode ser realizada em duas partes:

- a) por meio da inspeção visual do gráfico do parametro investigado contra o tempo; e
- b) através dos testes de constancia aplicados às séries formadas pelo defasamento da série de parametros (média, variancia); nesse caso para um parametro ser considerado periódico é suficiente que uma das séries defasadas não passe no teste.

Dois métodos (não-paramétrico e paramétrico) podem ser usados na tentativa de representar e remover a periodicidade em parametros e de se obter uma componente estocástica, estacionária de segunda ordem. Em ambos os métodos, a remoção da periodicidade na média e no desvio padrão consiste na transformação do tipo;

$$\epsilon_{p,\tau} = \frac{X_{p,\tau} - m_{\tau}}{s_{\tau}} \quad (\text{III.8})$$

onde m_{τ} é a média e s_{τ} é o desvio padrão.

A diferença entre os dois métodos está na forma de obtenção da componente periódica na média e no desvio padrão: no método não paramétrico elas são obtidas da própria série pelas equações:

$$m_{\tau} = \frac{1}{n} \sum_{p=1}^n x_{p,\tau} \quad (\text{III.9})$$

e,

$$s_{\tau} = \left\{ \frac{1}{n} \sum_{p=1}^n (x_{p,\tau} - m_{\tau})^2 \right\}^{1/2} \quad (\text{III.10})$$

onde n é o número total de anos, p é o p -ésimo ano e $\tau=1, 2, 3, \dots$, ω é a τ -ésima posição discreta do período base ω ; e no método paramétrico são ajustadas por funções periódicas (Análise de Fourier) com um número limitado de harmônicos, conforme descrito mais adiante.

A decisão quanto ao uso do método não paramétrico ou paramétrico está principalmente na economia do número de estatísticas necessárias para descrição da série. No primeiro caso requer o uso de 2ω estatísticas, ω de m_{τ} e ω de s_{τ} , o que implicaria para séries diárias num total de $2\omega = 730$ estatísticas, enquanto que no segundo caso apenas alguns harmônicos são necessários para descrever a periodicidade, geralmente menos de 20. No caso de 20 harmônicos, o total de estatísticas seria 82 que representa uma diminuição bastante significativa. Neste contexto, foi decidido utilizar o método paramétrico.

Os modelos que podem ser identificados quando se depara com a existência de periodicidade nos parâmetros de séries temporais geofísicas são do tipo:

TIPO	MÉDIA	VARIÂNCIA	COEFICIENTE DE VARIAÇÃO
1	periódica	constante	-
2	constante	periódica	-
3A	periódica	periódica	constante
3B	periódica	periódica	periódico

sendo, geralmente, o modelo 3B o mais adequado para séries meteorológicas.

Caso se empregue o modelo 3B e o método paramétrico, a componente estocástica ficaria determinada pelo uso da expressão (III.8) onde m_τ e s_τ são as funções periódicas ajustadas pela série de Fourier.

A componente estocástica obtida pelo método paramétrico requer, geralmente, uma padronização do tipo,

$$\epsilon_{p,\tau} = \frac{\epsilon_{p,\tau} - \mu_\epsilon}{\sigma_\epsilon} \quad (\text{III.11})$$

em que μ_ϵ é a média de $\epsilon_{p,\tau}$, estimada por $\bar{\epsilon}_{p,\tau}$ e σ_ϵ é o desvio padrão estimado por s_ϵ , tendo em vista que a mesma não possui exatamente média zero e variancia um.

3.3.2.1 - Análise de Fourier

A análise de Fourier é o método simples que permite colocar em evidencia as periodicidades presentes nas séries de tempo, e é o tipo de análise mais comumente a-

plicada à variações periódicas de séries meteorológicas. Este método, às vezes, vem a ser simplesmente o ajuste de uma função as séries analisadas.

As técnicas da análise por autocorrelação e da análise espectral também podem ser empregadas. Porém, no caso de séries meteorológicas, onde o período base de um ano é conhecido, estas técnicas fornecem informações de menor interesse do que a técnica da análise harmônica.

A periodicidade de uma série meteorológica pode apresentar-se na média, na variancia, no desvio padrão, nos coeficientes de correlação, nos momentos de ordens superiores ou nos parametros que são funções desses momentos. Então, pela análise de Fourier, os parametros periódicos das séries podem ser descritos por uma série de senos e co-senos (Série de Fourier), cuja equação geral é:

$$v_{\tau} = \bar{v} + \sum_{j=1}^m C_j \cos \left(\frac{2\pi j\tau}{\omega} + \theta_j \right) \quad (\text{III.12})$$

com v_{τ} representando o paramétrico periódico, \bar{v} a média do parametro v , j indica o número sequencial dos harmônicos no intervalo $j = 1, 2, \dots, \omega/2$, m é o número de harmônicos significantes, C_j é a amplitude do j -th harmônico e θ_j a fase do j -th harmônico, ω é o período.

A equação (III.12) pode ser escrita da forma:

$$v_{\tau} = \bar{v} + \sum_{j=1}^m (A_j \cos \frac{2\pi j \tau}{\omega} + B_j \sin \frac{2\pi j \tau}{\omega}) \quad (\text{III.13})$$

onde os coeficientes A_j e B_j são estimados por:

$$A_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} v_{\tau} \cos \frac{2\pi j \tau}{\omega} \quad (\text{III.14})$$

$$B_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} v_{\tau} \sin \frac{2\pi j \tau}{\omega} \quad (\text{III.15})$$

e

$$\bar{v} = \frac{1}{\omega} \sum_{\tau=1}^{\omega} v_{\tau} \quad (\text{III.16})$$

A amplitude do harmônico j é igual a

$$C_j = \sqrt{A_j^2 + B_j^2} \quad (\text{III.17})$$

e a fase

$$\theta_j = \text{tg}^{-1} (B_j/A_j) \quad (\text{III.18})$$

A variancia do harmônico j é dada por:

$$\text{var } h_j = \frac{C_j^2}{2} = \frac{A_j^2 + B_j^2}{2} \quad (\text{III.19})$$

Um plote de $C_j^2/2$ versus a frecuencia ordinária f_j ou a frecuencia angular λ_j , representa o periodograma ou espectro de variancia.

3.3.2.2 - Harmônicos significantes

O número máximo de harmônicos de uma função periódica v_T ajustadas a um parâmetro estatístico v é $\omega/2$ para períodos pares e $(\omega-1)/2$ para períodos ímpares, entretanto, comumente, poucos harmônicos são suficientes para descreverem a variação da função periódica.

A determinação do número de harmônicos significantes constitui, ainda, um problema no modelamento dos parâmetros periódicos. Os métodos analíticos para determinação dos harmônicos significantes sugeridos por Schuster (1898), Walker (1923) e Fisher (1929) apresentam limitações que os tornam inconvenientes para o tipo de fenômenos estudados.

Devido às dificuldades na aplicação destes testes tem sido desenvolvidos os seguintes métodos de investigação de harmônicos significantes:

1º) Método empírico de Yevjevich (1972a)

Este teste consiste na definição do parâmetro

$$\Delta p_j = \frac{\text{var } h_j}{s^2(v_T)} = \frac{C_j^2/2}{s^2(v_T)} = \frac{(A_j^2 + B_j^2)/2}{s^2(v_T)} \quad (\text{III.20})$$

que representa a parte da variância de v_T explicada pelo harmônico h_j . O somatório de Δp_j , ou seja

$$P_j = \sum_{i=1}^j \Delta p_j, \text{ para } j = 1, 2, \dots, m^* \quad (\text{III.21})$$

onde m^* é o número de harmônicos investigados, dá a parte da variação de v_T que é explicada pelos m^* harmônicos investigados.

Este método consiste em estabelecer uma faixa de aceitação da significancia dos harmônicos limitada por dois valores críticos,

$$P_{\min} = a \sqrt{\frac{\omega}{nc}} \quad \text{e} \quad P_{\max} = 1 - P_{\min} \quad (\text{III.22})$$

onde c é a ordem do momento usado para estimar o parametro v_T , ω e n são o período base e o número de anos de registro, e a é uma constante (usualmente $a=0.033$).

Os critérios usados para determinar os harmônicos significantes são:

- a) Se $P \leq P_{\min}$ não há harmônicos significantes na sequência dos valores v
- b) Se $P_{\min} < P \leq P_{\max}$ todos os harmônicos são considerados significantes
- c) Se $P > P_{\max}$ os primeiros j harmônicos, cujos P_j primeiros valores excedem P_{\max} são considerados significantes

29) Periodograma acumulado (Yevjevich, 1972a)

A razão entre a variancia acumulada dos primeiros m harmônicos e a variancia do estimado v_T de um parametro, dá o espectro linha acumulado

$$P_m = \frac{\sum_{j=1}^m \text{var } h_j}{\text{var } v_T} \quad (\text{III.23})$$

Este método é um processo gráfico baseado no conceito de que a variação de P_m como uma função de m , é composta por duas partes distintas:

1. parte periódica (rápido crescimento de P_m com m)
2. parte amostral (lento crescimento de P_m com m)

Dois caminhos podem ser tomados para verificação dos harmônicos significantes:

- a) As duas partes (periódica e amostral) são aproximadas por curvas suavizadas que se interceptam em um ponto. A frequência crítica desse ponto dá o número de harmônicos significantes, que são todos os harmônicos abaixo dessa frequência crítica.
- b) Assumir o modelo matemático aproximado das duas partes, estimar os parametros desse modelo, e achar a interseção das duas curvas. A frequência mais perto do ponto de interseção é a frequência crítica.

39) Método baseado na teoria dos intervalos de confiança, Sales (1977)

Este método consiste na determinação de intervalos, que, com uma probabilidade específica, contenham os valores do parametro periódico v_τ , e determinam, também, uma faixa de confiança para esse parametro no período fundamental ω .

O parametro v_τ é estimado com o número máximo de harmônicos e, com os harmônicos classificados em ordem de crescente de variancia, computa-se o parametro sucessivamente, pela subtração de um harmônico de cada vez, no sentido do menos para o mais significativo, até uma certa ordem, a partir da qual a função harmônica ajustada não esteja integralmente contida na faixa de confiança determinada. Em consequencia, para a ordem anterior a esta assim determinada, a função periódica ajustada estará contida na faixa de confiança. Esta ordem será considerada como limítrofe, sendo então o número de harmônicos para o parametro dado pela ordem anterior.

Os intervalos de confiança são dados por:

Para as médias:

Seja X_τ ($\tau=1, \dots, \omega$) uma variável aleatória com $N(\mu_\tau, \sigma_\tau^2)$.

No caso mais geral, se σ_τ é desconhecido, o intervalo de confiança para μ_τ é dado por:

$$P\left\{\bar{X}_\tau - t_{\lambda/2} \frac{s_\tau}{\sqrt{n}} \leq \mu_\tau \leq \bar{X}_\tau + t_{\lambda/2} \frac{s_\tau}{\sqrt{n}}\right\} = 1-\lambda \quad (\text{III.24})$$

onde, $t_{\lambda/2}$ é um valor da distribuição t de Student com $\gamma = n-1$ graus de liberdade, s_τ uma estimativa de σ_τ e λ o nível de significância.

O intervalo de confiança de $(1-\lambda)\%$ pode ser representado por,

$$\{C_{1\tau} \leq \mu_\tau \leq C_{2\tau}\}, \quad \tau = 1, 2, \dots, \omega \quad (\text{III.25})$$

Para as variancias:

O intervalo de confiança de $(1-\lambda)\%$ para a variancia é:

$$\frac{\sum_i (X_i - \mu_\tau)^2}{C_{2\tau}} \leq \sigma_\tau^2 \leq \frac{\sum_i (X_i - \mu_\tau)^2}{C_{1\tau}} \quad (\text{III.26})$$

onde $C_{1\tau}$ e $C_{2\tau}$ são os limites tais que a probabilidade de que um valor observado exceda $C_{1\tau}$ seja igual a $(1-\lambda/2)$ e de que não exceda $C_{2\tau}$ seja também igual a $(1-\tau/2)$, isto é:

$$P\{C_{1\tau} \leq \chi_{OBS}^2\} = 1 - \lambda/2$$

e

$$P\{\chi_{OBS}^2 \leq C_{2\tau}\} = 1 - \lambda/2$$

(III.27)

com $C_{1\tau}$ e $C_{2\tau}$ obtidos das tabelas de distribuição de χ^2 , com $\gamma = n$ graus de liberdade.

Para a média μ_τ desconhecida, o intervalo de confiança de $(1-\lambda)\%$ torna-se:

$$\frac{\sum_i (X_i - \bar{X}_\tau)^2}{C'_{2\tau}} \leq \sigma^2_\tau \leq \frac{\sum_i (X_i - \bar{X}_\tau)^2}{C'_{1\tau}} \quad (\text{III.28})$$

com os valores $C'_{1\tau}$ e $C'_{2\tau}$ obtidos das tabelas de distribuição de χ^2 , com $\gamma = n-1$ graus de liberdade.

3.3.3 - Componente estocástica

O conhecimento das características sequencias da componente estocástica é um passo fundamental na identificação do processo gerador. A identificação deste processo pode ser feita através do estudo da autocovariancia e do coeficiente de autocorrelação.

3.3.3.1 - Função de autocorrelação

O método de autocorrelação usa uma função particular chamada coeficiente de autocorrelação versus o defasamento entre os valores correlatados da sequencia, o que representado graficamente se chama correlograma. O comportamento do correlograma dá uma indicação sobre as propriedades de dependencia ou independencia entre os valores da série.

Os coeficientes de autocorrelação populacional são estimados pelos coeficientes de correlação serial r_k , através de duas aproximações: séries circulares e séries abertas, Yevjevich, (1972b). Em meteorologia, na aplicação prática, os valores de r_k podem ser estimados das amostras utilizando-se a aproximação de séries abertas, ou seja:

$$\rho_k = \frac{\text{cov}(\epsilon_i, \epsilon_{i+k})}{(\text{var } \epsilon_i \cdot \text{var } \epsilon_{i+k})^{1/2}} \quad (\text{III.29})$$

que é equivalente a

$$r_k = \frac{\frac{1}{N-k} \sum_{i=1}^{N-k} \epsilon_i \epsilon_{i+k} - \frac{1}{(N-k)^2} \left(\sum_{i=1}^{N-k} \epsilon_i \right) \left(\sum_{i=1}^{N-k} \epsilon_{i+k} \right)}{A \cdot B} \quad (\text{III.30})$$

sendo,

$$A = \left(\frac{1}{N-k} \sum_{i=1}^{N-k} \epsilon_i^2 - \frac{1}{(N-k)^2} \left(\sum_{i=1}^{N-k} \epsilon_i \right)^2 \right)^{1/2} \quad (\text{III.31})$$

$$B = \left(\frac{1}{N-k} \sum_{i=1}^{N-k} \epsilon_{i+k}^2 - \frac{1}{(N-k)^2} \left(\sum_{i=1}^{N-k} \epsilon_{i+k} \right)^2 \right)^{1/2} \quad (\text{III.32})$$

onde,

ϵ valores individuais da componente estocástica,

N tamanho da amostra, e

k defasamento em unidades de tempo.

3.3.3.2 - Modelo de dependencia da componente estocástica

A variável estocástica $\varepsilon_{p,\tau}$ obtida pela remoção da periodicidade na média e no desvio padrão da série $X_{p,\tau}$ (equação III.8 ou III.11) é aproximadamente uma série de tempo estacionária de 2a. ordem, se os coeficientes de au toregressão não são periódicos. O modelo de dependencia para $\varepsilon_{p,\tau}$ pode ser autoregressivo, do tipo médias móveis, uma combinação dos dois, ou de outros esquemas, mesmo de uma regressão não linear. Segundo Yevjevich (1972a), os modelos autoregressivos lineares são os que se aplicam me lhor aos fenômenos geofísicos. Neste trabalho serão in vestigados modelos autoregressivos lineares até sexta ordem.

O modelo autoregressivo geral de ordem m é:

$$\varepsilon_{p,\tau} = \sum_{k=1}^m \alpha_k \varepsilon_{p,\tau-k} + \sigma_{\zeta,\tau} \zeta_{p,\tau} \quad (\text{III.33})$$

onde,

α_k coeficientes autoregressivos como uma função do coeficiente de correlação serial $\rho_{k,\tau}$ (periódico ou não),

σ_{ζ} desvio padrão de $\zeta_{p,\tau}$ que é periódico se $\alpha_{k,\tau}$ for periódico, e

$\zeta_{p,\tau}$ componente estocástica independente padronizada

Se a $\zeta_{p,\tau}$ for, de fato independente, será aceita a hipótese de processo estocástico independente estacionário de 2a. ordem.

Para estacionaridade do processo é necessário que as raízes da equação característica do processo

$$1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_m B^m = 0 \quad (\text{III.34})$$

caiam fora do círculo unitário.

3.3.3.3 - Investigação da periodicidade nos coeficientes de autocorrelação e estimação dos coeficientes autoregressivos do modelo

O primeiro passo para a determinação do modelo de dependencia da componente estocástica consiste na pesquisa de existencia de periodicidade nos coeficientes de autocorrelação $\rho_{k,\tau}$ da série $\epsilon_{p,\tau}$.

Para verificar a existencia de periodicidade, Yevjevich (1972) propos a técnica de repartição da série (split sample) formado pelos coeficientes de autocorrelação da série $\epsilon_{p,\tau}$. Esta série é estimada pelos coeficientes de correlação serial $r_{k,\tau}$ da seguinte forma:

$$r_{k,\tau} = \frac{\sum_{p=1}^N (\epsilon_{p,\tau} - \bar{\epsilon}_{p,\tau}) (\epsilon_{p,\tau+k} - \bar{\epsilon}_{p,\tau+k})}{\left\{ \sum_{p=1}^N (\epsilon_{p,\tau} - \bar{\epsilon}_{p,\tau})^2 \sum_{p=1}^N (\epsilon_{p,\tau+k} - \bar{\epsilon}_{p,\tau+k})^2 \right\}^{\frac{1}{2}}}$$

(III.35)

para $\tau=1, 2, \dots, \omega$ e $k=1, 2, \dots, N$ sendo N o número de anos de registro e k o defasamento. Quando o índice τ é igual a $(\omega-k+1)$, N deverá ser substituído por $(N-1)$ e $\epsilon_{p,\tau+k}$ por $\epsilon_{p+1,k}$.

A técnica de repartição da série $r_{k,\tau}$ visa eliminar o aparecimento de correlação espúria proveniente da dependencia entre os sucessivos valores de $r_{k,\tau}$, introduzida pela maneira de computá-los. Nesta técnica, a formação das subséries se faz do seguinte modo: Para um dado k , formam-se duas subséries, sendo que a primeira subsérie consiste dos primeiros k valores da sequencia, enquanto os próximos valores de k são parte da segunda subsérie, alternando até abranger todos os valores de $r_{k,\tau}$ ou seja:

Para $k=1$ teríamos duas subséries, uma formada pelos índices impares e outra formada por todos os índices pares.

Para $k=2$ a primeira subsérie seria, $r_{2,1}, r_{2,2}, r_{2,5}, r_{2,6}, r_{2,9}, r_{2,10}, \dots$, e, a segunda subsérie seria $r_{2,3}, r_{2,4}, r_{2,7}, r_{2,8}, r_{2,11}, r_{2,12}$ e assim seguidamente

Para $k=3$ a primeira subsérie seria $r_{3,1}, r_{3,2}, r_{3,3}, r_{3,7}, r_{3,8}, r_{3,9}, r_{3,13}, \dots$ e, a segunda subsérie seria $r_{3,4}, r_{3,5}, r_{3,6}, r_{3,10}, r_{3,11}, r_{3,12}, r_{3,16}, \dots$, e assim por diante.

As subséries assim constituídas podem ser testadas quanto a periodicidade através dos testes de constância de parâmetros usando o teste de Student e Fisher ou pelo uso da análise harmônica, investigando os harmônicos em ambas subséries. Neste último caso, a periodicidade é aceita somente nos casos em que existam o mesmo número de harmônicos significantes em ambas as subséries.

Não existindo periodicidade nos coeficientes de autocorrelação, os coeficientes autoregressivos do modelo (α_k para $k=1,2, \dots, m$, onde m é a ordem do modelo) podem ser estimados através das equações de Yule-Walker, (Box e Jenkins, 1970);

$$\begin{aligned}
 \rho_1 &= \alpha_1 + \alpha_2 \rho_1 + \dots + \alpha_m \rho_{m-1} \\
 \rho_2 &= \alpha_1 \rho_1 + \alpha_2 + \dots + \alpha_m \rho_{m-2} \\
 &\vdots \\
 \rho_m &= \alpha_1 \rho_{m-1} + \alpha_2 \rho_{m-2} + \dots + \alpha_m
 \end{aligned}
 \tag{III.36}$$

que tem a seguinte forma matricial:

$$\begin{bmatrix}
 1 & \rho_1 & \rho_{m-1} & \alpha_1 \\
 \rho_1 & 1 & \rho_{m-2} & \alpha_2 \\
 \vdots & \vdots & \vdots & \vdots \\
 \rho_{m-1} & \rho_{m-2} & 1 & \alpha_m
 \end{bmatrix} = \begin{matrix} \rho_1 \\ \rho_2 \\ \vdots \\ \alpha_m \end{matrix} \quad (\text{III.37})$$

Esta matriz é resolvida substituindo-se o coeficiente de autocorrelação ρ_m pelo seu estimado r_m .

Os valores dos coeficientes de correlação serial r_m para estimação de α_m , podem ser empregados da seguinte forma:

- 1- Usando as médias dos valores $r_{1,\tau}, r_{2,\tau}, \dots, r_{k,\tau}$ da série $\epsilon_{p,\tau}$ designadas por r_1, r_2, \dots, r_k , (equação III.35).
- 2- Usando os valores $r_{1,\tau}, r_{2,\tau}, \dots, r_{k,\tau}$ do correlograma sobre toda a série $\epsilon_{p,\tau}$ ou seja, sobre todos os N valores da série $\epsilon_{p,\tau}$ (equação III.30).

Os valores computados para os dois casos diferem frequentemente muito pouco.

Quando ρ_k é periódico, α_k também é periódico. Neste caso ρ_k é estimado por $r_{k,\tau}$ utilizando-se a equação (III.13) com v_τ substituído por $r_{k,\tau}$ para $k=1,2, \dots$ e os coeficientes de autoregressão $\alpha_{k,\tau}$, são obtidos através das equações de Yule-Walker devidamente modificadas.

3.3.3.4 - Desvio padrão de $\zeta_{p,\tau}$

O desvio padrão da componente estocástica independente obtida pelo modelo autoregressivo até 6a. ordem, é:

1a. ordem:

$$\sigma_{\zeta} = (1 + \alpha_1^2 - 2\alpha_1 r_1)^{1/2} \quad (\text{III.38})$$

2a. ordem:

$$\sigma_{\zeta} = (1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1 r_1 - 2\alpha_2 r_2 + 2\alpha_1 \alpha_2 r_1)^{1/2} \quad (\text{III.39})$$

3a. ordem:

$$\begin{aligned} \sigma_{\zeta} = (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 - 2\alpha_1 r_1 - 2\alpha_2 r_2 - 2\alpha_3 r_3 + \\ + 2\alpha_1 \alpha_2 r_1 + 2\alpha_1 \alpha_3 r_2 + 2\alpha_2 \alpha_3 r_1)^{1/2} \end{aligned} \quad (\text{III.40})$$

4a. ordem:

$$\begin{aligned} \sigma_{\zeta} = (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 - 2\alpha_1 r_1 - 2\alpha_2 r_2 - 2\alpha_3 r_3 - \\ - 2\alpha_4 r_4 + 2\alpha_1 \alpha_2 r_1 + 2\alpha_1 \alpha_3 r_2 + 2\alpha_1 \alpha_4 r_3 + \\ + 2\alpha_2 \alpha_3 r_1 + 2\alpha_2 \alpha_4 r_2 + 2\alpha_3 \alpha_4 r_1)^{1/2} \end{aligned} \quad (\text{III.41})$$

5a. ordem:

$$\begin{aligned}
 \sigma_{\zeta} = & (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 + \alpha_5^2 - 2\alpha_1 r_1 - 2\alpha_2 r_2 - \\
 & - 2\alpha_3 r_3 - 2\alpha_4 r_4 - 2\alpha_5 r_5 + 2\alpha_1 \alpha_2 r_1 + 2\alpha_1 \alpha_3 r_2 + \\
 & + 2\alpha_1 \alpha_4 r_3 + 2\alpha_1 \alpha_5 r_4 + 2\alpha_2 \alpha_3 r_1 + 2\alpha_2 \alpha_4 r_2 + \\
 & + 2\alpha_2 \alpha_5 r_3 + 2\alpha_3 \alpha_4 r_1 + 2\alpha_3 \alpha_5 r_2 + 2\alpha_4 \alpha_5 r_1)^{1/2}
 \end{aligned}$$

(III.42)

6a. ordem:

$$\begin{aligned}
 \sigma_{\zeta} = & (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 + \alpha_5^2 + \alpha_6^2 - 2\alpha_1 r_1 - 2\alpha_2 r_2 - \\
 & - 2\alpha_3 r_3 - 2\alpha_4 r_4 - 2\alpha_5 r_5 - 2\alpha_6 r_6 + 2\alpha_1 \alpha_2 r_1 + \\
 & + 2\alpha_1 \alpha_3 r_2 + 2\alpha_1 \alpha_4 r_3 + 2\alpha_1 \alpha_5 r_4 + 2\alpha_1 \alpha_6 r_5 + \\
 & + 2\alpha_2 \alpha_3 r_1 + 2\alpha_2 \alpha_4 r_2 + 2\alpha_2 \alpha_5 r_3 + 2\alpha_2 \alpha_6 r_4 + \\
 & + 2\alpha_3 \alpha_4 r_1 + 2\alpha_3 \alpha_5 r_2 + 2\alpha_3 \alpha_6 r_3 + 2\alpha_4 \alpha_5 r_1 + \\
 & + 2\alpha_4 \alpha_6 r_2 + 2\alpha_5 \alpha_6 r_1)^{1/2}
 \end{aligned}$$

(III.43)

3.3.3.5 - Seleção da ordem do modelo autoregressivo

Vários métodos são encontrados na literatura para a determinação da ordem apropriada dos modelos autore-

gressivos, entre eles:

1. Uso do espectro de densidade de variancia. Um exame visual da forma do espectro pode indicar a ordem do processo autoregressivo.
2. Obtenção da componente estocástica ϵ_T , presumidamente independente de um modelo autoregressivo de ordem m e testar sua independencia. Se a série ζ_T é uma série aleatória independente, o modelo é considerado como um modelo de dependencia da série ϵ_T .
3. Teste de Quenoille (1949) para grandes amostras. Esta técnica é um modelo laborioso que inclui computação de dois conjuntos de constantes e um teste paramétrico.
4. Método simplificado apresentado por Yevjevich (1972a) A medida do bom ajuste do modelo usado neste método é expressa pela determinação dos coeficientes D_i , $i=1, 2, 3, \dots$ que dão a percentagem da variancia total de $\epsilon_{p,T}$ explicada pelo i -ésimo termo da equação autoregressiva, enquanto a parte restante da variancia de $\epsilon_{p,T}$ é explicada pelo termo $(\sigma_{p,\zeta} \epsilon_{p,\zeta})$.

Os coeficientes de determinação dos modelos autoregressivos de 1a. a 3a. ordem são dados respectivamente por:

$$D_1 = r_1^2 \quad (\text{III.44})$$

$$D_2 = \frac{r_1^2 + r_2^2 - 2r_1^2 r_2}{1 - r_1^2} \quad (\text{III.45})$$

$$D_3 = \frac{r_1^2 + r_2^2 + r_3^2 + 2r_1^2 r_3 + 2r_1^2 r_2^2 + 2r_1 r_2^2 r_3 - 2r_1^2 r_2}{1 - 2r_1^2 - r_2^2 + 2r_1^2 r_2} - \frac{4r_1 r_2 r_3 - r_1^4 - r_2^4 - r_1^2 r_3^2}{1 - 2r_1^2 - r_2^2 + 2r_1^2 r_2} \quad (\text{III.46})$$

em que r_1 , r_2 e r_3 podem ser os valores $\bar{r}_{k,\tau}$ ou os valores $r_{1,\varepsilon}$, $r_{2,\varepsilon}$, $r_{3,\varepsilon}$ conforme item 3.3.3.3.

O critério de seleção do modelo é baseado na comparação dos coeficientes de determinação. Se $D_2 - D_1 \leq 0.01$, o modelo de 1a. ordem é aceito. Se $D_2 - D_1 > 0.01$, mas $D_3 - D_1 \leq 0.01$, $D_3 - D_2 > 0.01$ o modelo de 3a. ordem será aceito.

Qualquer valor ΔD_i pequeno, ou seja, 0.01, 0.02 ou similar não causa diferença significativa nos resultados. Para ordens superiores a três os coeficientes de determinação, D_i , podem ser determinados pela diferença $(1 - \sigma^2)$, onde σ^2 é a variancia da componente estocástica independente, que dá a variancia explicada pelos termos

autoregressivos. O critério para determinar a ordem do modelo poderia ser $D_i - D_{i-1} \leq e$ com $e = 0.01$ ou um número similar.

3.3.3.6 - Teste de adequacidade dos modelos autoregressivos

A adequacidade dos modelos autoregressivos pode ser pesquisada através de testes que verificam a independência da componente estocástica $\zeta_{p,\tau}$, presumidamente independente, obtida pelo uso de esquemas autoregressivos.

Anderson (1941) propôs um teste que fixa dois limites para testar diferenças significantes do valor esperado do coeficiente de correlação para um dado nível de significancia. Este teste foi desenvolvido para séries circulares, mas pode ser usado para séries abertas se a amostra é suficientemente grande como é o caso de séries diárias, Tao e outros (1976). Usualmente, o primeiro coeficiente de correlação é de maior importancia. Os limites de tolerancia para o nível de significancia de 95% são:

$$\frac{-1 + 1.96 \sqrt{N-3}}{N-2} \quad \text{e} \quad \frac{-1 - 1.96 \sqrt{N-3}}{N-2} \quad (\text{III.47})$$

com N sendo o número de observações.

Siddiqui (1957) aplicando uma aproximação para o caso de séries abertas, apresentou a distribuição de r_1 para variáveis normais independentes:

$$f(r_1) = \frac{(1 - r_1)^{\frac{N-1}{2}} (1 + r_1)^{\frac{N-3}{2}}}{2^{N-1} B(\frac{N+1}{2}, \frac{N-1}{2})} + O(\frac{1}{N^2}) \quad (\text{III.48})$$

sendo N o tamanho da amostra, $B(\frac{N+1}{2}, \frac{N-1}{2})$ a função Beta e o termo $O(1/N^2)$ uma quantidade pequena que decresce rapidamente quanto N aumenta.

A média e a variancia de r_1 , são:

$$E(r_1) = -\frac{1}{N} \quad (\text{III.49})$$

$$\text{var}(r_1) = \frac{N^3 - 3N^2 + 4}{N^2(N^2 - 1)} \quad (\text{III.50})$$

Para $N \geq 30$, a distribuição de r_1 aproxima-se de uma distribuição normal. Quando N é muito grande e k é muito pequeno, a equação (III.45) pode ser aplicada à distribuição de r_k . Os limites de tolerancia para r_k para o nível de significancia de 95%, são:

$$-\frac{1}{N} \pm 1.96 \sqrt{\frac{N^3 - N^2 + 4}{N^2(N^2 - 1)}} \quad (\text{III.51})$$

3.3.4 - Funções de distribuição de probabilidade para ajuste da componente estocástica independente

Obtida a componente estocástica independente ou aleatória de um fenômeno, torna-se de grande valia formular um modelo matemático que ajudará a investigar, de maneira precisa, esse fenômeno. A formulação do modelo é feita através do ajuste da componente independente à uma distribuição de probabilidade teórica.

A seguir serão apresentadas algumas distribuições de probabilidade teóricas que comumente ajustam-se à distribuição de frequências observadas da componente estocástica independente, de acordo com Tao e outros, 1976.

3.3.4.1 - Função de distribuição de probabilidade normal

Esta distribuição simétrica tem a seguinte expressão:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty \leq x \leq \infty \quad (\text{III.52})$$

com x a variável, μ a média da população, e σ o desvio padrão da população. μ e σ são estimados pelo método de máxima verossimilhança, com N o tamanho da amostra, ou seja:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

e

$$\hat{\sigma} = \left\{ \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right\}^{1/2}$$

3.3.4.2 - Função de distribuição de probabilidade lognor- mal - 3 parâmetros

A equação que define esta função é:

$$f(x) = \frac{1}{(x-\beta)\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x-\beta) - \ln \mu)^2}{2\sigma^2}}, \quad \beta \leq x \leq \infty$$

(III.53)

com μ a média geométrica da população de $(x-\beta)$, σ o desvio padrão da população de $\ln(x-\beta)$, e β o limite inferior do intervalo de definição. β é estimado por $\hat{\beta}$ pelo método de máxima verossimilhança através da resolução, por um processo iterativo, da seguinte equação:

$$\sum_{i=1}^N \frac{1}{x_i - \beta} \left\{ \frac{1}{N} \sum_{i=1}^N \ln^2(x_i - \beta) - \left(\frac{1}{N} \sum_{i=1}^N \ln(x_i - \beta) \right)^2 - \right.$$

$$\left. - \frac{1}{N} \sum_{i=1}^N \ln(x - \beta) \right\} + \sum_{i=1}^N \frac{\ln(x_i - \beta)}{(x_i - \beta)} = 0 \quad (\text{III.54})$$

Os parametros estimados desta distribuição por máxima verossimilhança são:

$$\widehat{\ln \mu} = \frac{1}{N} \sum_{i=1}^N \ln(x_i - \beta)$$

e

(III.55)

$$\hat{\sigma} = \left\{ \frac{1}{N} \sum_{i=1}^N (\ln(x_i - \beta) - \widehat{\ln \mu})^2 \right\}^{1/2}$$

3.3.4.3 - Função de distribuição de probabilidade gama-3 parametros ou Pearson Tipo III

A expressão que define esta função é

$$f(x) = \frac{1}{\beta \Gamma(\alpha)} \frac{(x-\gamma)^{\alpha-1}}{\beta} e^{-\frac{x-\gamma}{\beta}} \quad \beta \leq x \leq \infty \quad (III.56)$$

sendo α o parametro de forma, β o parametro de escala, γ o parametro de locação do limite inferior, e $\Gamma(\alpha)$ a função gama de α . γ é estimado por máxima verossimilhança pela resolução por procedimento iterativo da seguinte equação:

$$\frac{1 + (1 + \frac{4}{3}A)^{1/2}}{1 + (1 + \frac{4}{3}A)^{1/2} - 4A} - (\bar{x} - \bar{\gamma}) \frac{1}{N} \sum_{i=1}^N \frac{1}{x_i - \bar{\gamma}} = 0 \quad (III.57)$$

em que

$$A = \ln(\bar{x} - \bar{\gamma}) - \frac{1}{N} \sum_{i=1}^N \ln(x_i - \bar{\gamma}) \quad (III.58)$$

com \bar{x} a média dos N valores de x . Uma vez que γ seja determinado, o parametro α é estimado por

$$\hat{\alpha} = \frac{(1 + \frac{4}{3}A)^{1/2}}{4A} - \Delta\alpha \quad (III.59)$$

com A dado pela equação (III.58) e $\Delta\alpha$ aproximado por

$$\Delta\alpha = 0.04475 (0.26)^\alpha \quad (\text{III.60})$$

o parametro β é então estimado por

$$\beta = \frac{1}{\hat{\alpha}} (\bar{x} - \hat{\gamma}) \quad (\text{III.61})$$

3.3.4.4 - Função densidade de probabilidade "Double-branch gamma"

Esta função é definida por

$$f(x) = \frac{P}{\beta_1^{\alpha_1} \Gamma(\alpha_1)} (x_0 - x)^{\alpha_1 - 1} e^{-\frac{x_0 - x}{\beta_1}} + \quad (-\infty, x_0) \quad (\text{III.62})$$

$$+ \frac{1-P}{\beta_2^{\alpha_2} \Gamma(\alpha_2)} (x - x_0)^{\alpha_2 - 1} e^{-\frac{x - x_0}{\beta_2}} \quad (x_0, \infty)$$

com P definido como $P = P(x \leq x_0)$ e $1-P = P(x > x_0)$, x_0 a moda, α_1 e β_1 parametros do ramo esquerdo, e α_2 e β_2 parametros do ramo direito. A moda x_0 é melhor estimada pela seleção do intervalo de classe Δx , da ordem de 0.001 ou menor, e é computada por $\hat{x}_0 = x_L + \Delta x/2$, em que x_L é o limite inferior do intervalo Δx com o maior número de frequencia. O parametro P é determinado por $P = n_1/N$, com n_1 o número de valores de x satisfazendo a condição $x_1 \leq x_0$. Os parametros α_1 e β_1 do ramo esquerdo são estimados pelo uso somente dos n_1 valores com $x_1 \leq x_0$, enquanto α_2 e β_2 do ramo direito são estimados

pelos n_2 valores amostrais restantes sendo $n_2 = N - n_1$, para $x_i > x_0$. Os parâmetros α_1 e β_1 são estimados por:

$$\hat{\alpha}_1 = \frac{1 + \sqrt{1 + \frac{4}{3} A_1}}{4A_1} = 0.04475 (0.26)^{\alpha_1} \quad (\text{III.63})$$

com

$$A_1 = \ln \left\{ \hat{x}_0 - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \right\} - \frac{1}{n_1} \sum_{i=1}^{n_1} \ln (\hat{x}_0 - x_i) \quad (\text{III.64})$$

e

$$\beta_1 = \frac{1}{\hat{\alpha}_1} \left(\hat{x}_0 - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \right) \quad (\text{III.65})$$

Similarmente, os parâmetros α_2 e β_2 são estimados pelo uso das equações (III.63 a III.65) com o termo $x_0 - x_i$ substituído por $x_i - x_0$ e n_1 e n_2 . Os parâmetros de forma α_1 e β_2 para os dois ramos serão iguais ou menores que 1. Se a equação (III.63) dá valores maiores que a unidade, a função gama de um pico será substituída pela função exponencial-J.

3.3.4.5 - Teste para verificação de ajuste da distribuição de frequência

Vários métodos podem ser usados para testar o bom ajuste de uma função de distribuição de probabilidade à distribuição de frequências observadas.

Esses métodos são o teste de qui-quadrado, o teste de razão de máxima verossimilhança, e o teste de Smirnov-Kolmogorov. O teste de qui-quadrado foi selecionado neste trabalho por ser bastante conhecido e aplicado em estatística e meteorologia.

As etapas básicas do teste de qui-quadrado são sumarizadas como segue, Markovic (1965). O total de observações é dividido em k mutuamente exclusivos e exaustivos intervalos de classe, de igual probabilidade ou igual amplitude, tendo cada intervalo a probabilidade de classe observada O_j e a probabilidade de classe correspondente esperada E_j ($j=1,2,\dots,k$). O valor esperado E_j é usado como padrão da classe e a quantidade $(O_j - E_j)^2 / E_j$ é a medida do afastamento do padrão. A medida da discrepância total, χ^2 , entre as frequências observadas e a esperada é então

$$\hat{\alpha}_1 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} \quad (\text{III.66})$$

A distribuição amostral de χ^2 segue, com muita aproximação, a distribuição de qui-quadrado com γ graus de liberdade quando as frequências esperadas forem, pelo menos, iguais a 5, melhorando as aproximações para valores maiores. O número de graus de liberdade γ é dado por:

- a) $\gamma = k - 1$, quando as frequências esperadas puderem ser calculadas, sem que se faça estimativas dos parâmetros populacionais, a partir de estatísticas amostrais.
- b) $\gamma = k - 1 - m$, quando as frequências esperadas somente puderem ser calculadas mediante a estimativa de m parâmetros populacionais, a partir de estatísticas amostrais.

A seleção do número de intervalos de classe, k , para a aplicação do teste de qui-quadrado foi proposta por Sturges (1926) através da expressão empírica:

$$k = 1 + 3,3 \ln N \quad (\text{III.67})$$

com N sendo o tamanho da amostra.

IV - APLICAÇÕES

4.1 - Séries empregadas e análise preliminar

4.1.1. - Introdução

Fazendo o levantamento dos registros de radiação solar global e temperatura do ar constatou-se a pouquíssima disponibilidade dos registros de radiação solar global no Brasil, o que impediu a concretização da terceira meta do trabalho, ou seja: obter um modelo estocástico para estimar a radiação solar em função da temperatura média do ar. Em presença dessa impossibilidade, serão somente cumpridas as duas primeiras etapas do objetivo deste trabalho.

O material utilizado neste trabalho são as séries diárias de radiação solar global e de temperatura média diária referente as estações meteorológicas Ecologia Agrícola situada na Universidade Rural - Km 47 - Latitude $22^{\circ} 46' S$ e longitude $43^{\circ} 41' W$ (período 1963 a 1973) e Estação Centro Experimental de Campinas situada no Instituto Agronômico de Campinas - Latitude $22^{\circ} 05' S$ e longitude $47^{\circ} 05' W$ (período 1965 a 1974).

O aparelho registrador da radiação solar global para ambas as localidades é um actinógrafo bimetálico de fabricação Fuess. As temperaturas médias diárias foram obtidas das observações dos termômetros de máxima, mínima e ordinário, de fabricação Fuess.

4.1.2 - Avaliação dos dados disponíveis

A. Radiação Solar Global

Os dados de radiação solar global da Estação Centro Experimental de Campinas foram processados pelo Instituto Agrônomo de Campinas. O processo usado para cotação dos diagramas do actinôgrafo conforme informação do referido Instituto foi o processo de cotação horária. Este processo consiste em dividir o diagrama no eixo do tempo em intervalos de uma hora. A partir dessa divisão encontra-se a altura do ponto médio do intervalo e em seguida obtem-se a área abaixo da curva registrada, conforme a equação.

$$A = \sum (\text{altura} \times \Delta t) \quad (\text{IV.1})$$

onde, altura é a altura do ponto médio e Δt é o intervalo de tempo (uma hora).

A área encontrada, deverá ser convertida em unidades do gráfico e posteriormente multiplicada pela constante de calibração do instrumento.

Os dados de radiação solar global diária da Estação Ecologia Agrícola foram levantados a partir dos diagramas do actinôgrafo utilizando-se um planímetro de compensação marca KOIZUMI - Tipo KP-27. Este processo de leitura dos diagramas fornece bastante precisão desde que sejam obedecidas as regras de uso do planímetro. Foram realizadas três leituras de cada diagrama e em seguida obtida a média aritmética. Essa média após multiplicada pela constante do planímetro fornecerá a área

sob a curva registrada. Usando o mesmo processo anterior, ou seja, a constante do gráfico e constante de calibração do instrumento, obteremos a radiação solar global diária.

Como não foi possível encontrar a constante da calibração do instrumento para este posto foi assumido para a mesma o valor 1 (um). Este fato deve ser tomado em conta quando se tentar a aplicação dos resultados encontrados.

B - Temperatura média diária

As temperatura médias diárias para ambas as estações foram obtidas a partir da leituras dos termômetros de máxima, de mínima e ordinário empregando-se a seguinte fórmula, Serra (1963):

$$TMD = \frac{T_{\max} + T_{\min} + T_{12} + 2 \times T_{24}}{5} \quad (IV.2)$$

onde os índices 12 e 24 representam a temperatura do ar às 12 e 24 horas de Greenwich, respectivamente.

4.1.3 - Preenchimento das falhas

As falhas existentes nos dados de radiação solar global estão apresentados na tabela (IV.1). Desta tabela pode ser observado que a porcentagem de dias sem dados é 3.2% para a Estação Campo Experimental de Campinas e 7% para a Estação Ecologia Agrícola.

A série de temperatura média diária da Estação Experimental de Campinas não apresentou nenhuma falha e a Estação Ecologia Agrícola as seguintes:

1964 - Novembro e dezembro, totalizando 61 dias

1965 - 15 dias em maio e 13 em junho, totalizando 23 dias

O número de falhas total desta série é portanto 89 dias para um período de 11 anos que representa 2.2% da série toda.

Para a análise a ser efetuada nestas séries, é necessário que as séries estejam completas. Para preenchimento das falhas foi empregada a Subrotina GAUSS do SSP (Subroutine Scientific Package, IBM) implantada no sistema Burroughs-6700 instalado no Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro. Esta subrotina computa números aleatórios normalmente distribuídos com uma dada média e desvio padrão. Cada falha foi preenchida empregando-se as médias e desvios padrões diários, obtidos com os N (nº de anos) valores diários da amostra. Aqui cabe assinalar a incompatibilidade entre o método escolhido para o preenchimento das falhas e a hipótese de dependência entre os valores das séries na análise estrutural. Entretanto, como o número de falhas representa uma porcentagem pequena da amostra, e os métodos alternativos de preenchimento de falhas (por exemplo usando a média diária de longo período) não oferecem vantagens sob o método selecionado, é considerado o preenchimento feito como adequado para os objetivos do estudo. A alternativa de se utilizar um posto vizinho para correlações foi impossível tendo em vista a carência de postos.

TABELA IV.1 - Número de falhas existentes nos dados diários de Radiação Solar Global nos postos: (1) Estação Centro Experimental de Campinas, (2) Estação Ecologia Agrícola, (x) dados não disponíveis e (-) nenhuma falha existente.

RADIÇÃO SOLAR													
ANO	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	TOTAL
POSTO	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2
JAN.	x 4	x 1	- -	5 5	- 5	2 -	6 -	- -	4 -	- 2	- -	- x	17 17
FEV.	x 8	x 6	28 -	- 1	- 3	- -	- 3	- -	1 -	- -	- -	- x	29 21
MAR.	x -	x 5	7 -	3 -	- 2	3 -	- 7	- 4	- -	- -	- 2	- x	13 20
ABR.	x 1	x 2	- -	- 12	- -	- 6	1 5	- -	- -	- -	- 1	- x	1 27
MAI.	x -	x -	- -	- -	- 2	- 5	- 5	- -	- -	- -	- 10	- x	- 22
JUN.	x 6	x 4	- -	- -	- 2	- -	- 4	- -	- -	- -	- 6	- x	- 22
JUL.	x 3	x -	- -	5 -	- 4	- -	1 -	- -	- -	- 1	- 2	- x	6 10
AGO.	x 1	x 9	- -	4 -	- -	1 4	- 4	1 1	- 6	- -	- 1	- x	6 26
SET.	x 1	x -	- 9	- 8	- -	- 2	- 5	4 -	- -	- -	- 2	- x	4 27
OUT.	x -	x -	- 1	- 1	3 -	2 1	- 6	1 -	- -	- 1	- 1	- x	6 11
NOV.	x 5	x -	- 3	- 21	- -	- 3	- -	5 1	10 -	- -	- 1	- x	15 34
DEZ.	x 2	x -	- 14	- 13	1 2	2 -	- 4	11 1	- 2	- 2	5 6	- x	19 46
TOTAL	x 31	x 27	35 27	17 61	4 20	10 21	8 43	22 7	15 8	- 6	5 32	- x	116 283

4.2 - Emprego da Metodologia

4.2.1 - Verificação da homogeneidade nas séries

Para verificação da homogeneidade nas séries foram aplicados os testes t de Student e F de Fisher conforme apresentado no item 3.3.1 utilizando-se as equações III.5 e III.6 respectivamente. O nível de significância selecionado foi de 5% e os testes foram aplicados às séries formadas pelos valores anuais obtidos por média aritmética dos valores diários que estão plotados na figura (IV.2 a IV.9). Os resultados encontram-se na tabela (IV.2), e desta pode ser verificado que somente a série de radiação solar global da Estação Centro Experimental de Campinas não passou no teste de constância das médias. Pela inspeção visual desta série anual, figura (IV. 1 (a)), pode ser observado uma mudança em seu comportamento a partir de 1969. Como o número de anos observados é muito pequeno (10 anos) e não se tem a certeza de uma causa para esta mudança, não seria recomendável considerar a remoção desta possível tendência, porém deve ser examinado mais exhaustivamente o processo de coleta e processamento inicial dos dados.

ESTAÇÃO	VARIÁVEL	N ₁ (anos)	N ₂ (anos)	TESTE DE IGUALDADE DE MÉDIAS (STUDENT)		TESTE DE IGUALDADE DE VARIÂNCIAS (FISHER)	
				t crítico (5%)	t calcula do	F crítico (5%)	t calcula do
CAMPINAS	RADIAÇÃO	5	5	2.306	3.820	6.39	2.64
	TEMPERATURA	5	5	2.306	0.764	6.39	3.66
ECOLOGIA AGRÍCOLA	RADIAÇÃO	5	6	2.262	0.020	6.26	2.04
	TEMPERATURA	5	6	2.262	0.656	6.26	5.10

TABELA IV.2 - Verificação da homogeneidade das séries anuais de radiação solar global e temperatura média do ar com os Testes de Student e Fisher. (N₁ e N₂ tamanho das sub-amostras).

4.2.2 - Parâmetros estatísticos das amostras

Os parâmetros estatísticos das amostras foram determinados conforme as expressões apresentadas no anexo I. As médias diárias de radiação solar global e de temperatura do ar estão plotadas nas figuras (IV.2 a 5). Os desvios padrões diários da radiação solar global e temperatura do ar nas figuras (IV.6 a 9). A tabela (IV.3) apresenta os parâmetros estatísticos das amostras.

PARÂMETROS	ESTAÇÃO DE CAMPINAS		ESTAÇÃO ECOLOGIA AGRÍCOLA	
	RADIAÇÃO	TEMPERATURA	RADIAÇÃO	TEMPERATURA
Média	370.1	20.6	402.7	23.1
Variância	13310.2	9.1	24434.2	9.7
Desvio Padrão	115.4	3.0	156.3	3.2
Coef. de Assi met.	- 0.16	- 0.4	- 0.04	- 0.03
Coef.de Curtose	2.9	3.1	2.6	2.4
Coef.de Variação	0.31	0.15	0.38	0.13
Média das Mé dias Diárias	370.1	20.6	402.7	23.1
Média das Va riâncias Diá rias	9615.1	4.8	19204.3	4.9

TABELA IV.3 - PARÂMETROS ESTATÍSTICOS DAS AMOSTRAS

4.2.3 - Componente periódica

A pesquisa sobre a existência da periodicidade na média, desvio padrão e coeficiente de variação das amostras foram realizadas conforme item (3.3.2). O caráter periódico dos mesmos pode ser visualizado nas Figuras (IV.2 a IV.9) e observado quantitativamente através dos resultados obtidos com a aplicação dos Testes de Student e Fisher (item 3.3.1) ao nível de significância de 5% de acordo com o seguinte procedimento: Construí-se doze subséries de seis meses cada, e em seguida foram aplicados os testes mencionados para cada par de subséries, a saber:

1º Par: Jan a Jun e Jul a Dez
 2º Par: Fev a Jul e Ago a Jan
 3º Par: Mar a Ago e Set a Fev
 4º Par: Abr a Set e Out a Mar
 5º Par: Mai a Out e Nov a Abr
 6º Par: Jun a Nov e Dez a Mai

Os resultados estão apresentados na tabela (IV.4).

Nesta tabela observa-se que a estatística t e F calculada para todas as subséries apresentam valores maiores que os valores críticos de t e F , o que sugere a existência de periodicidade.

O método paramétrico foi utilizado para remover a componente periódica na média e variância. A componente periódica na média e variância foram ajustadas utilizando-se a equação (III.13). Para determinação do número de harmônicas significantes, foram computadas as 182 harmônicas possíveis de cada série e aplicados os 3 testes descritos no item (3.3.2).

A seguir serão apresentados os resultados dos três testes aplicados:

a) Teste Empírico de Yevjevich

Os resultados desse teste estão apresentados na tabela (IV.5).

Pela aplicação dos limites $P_{\text{máx}}$ e $P_{\text{mín}}$ nos resultados apresentados na tabela (IV.5) fica difícil determinar o número de harmônicos significativos, com excessão do parâmetro média das séries de temperatura.

PARÂMETROS	PARES DE SUB-SÉRIES	CAMPINAS				ECOLOGIA AGRÍCOLA			
		RADIAÇÃO		TEMPERATURA		RADIAÇÃO		TEMPERATURA	
		S	F	S	F	S	F	S	F
MÉDIA	1	0.39	1.29	3.62	1.49	3.75	1.07	7.09	1.40
	2	7.75	1.43	4.94	2.20	4.34	1.16	2.15	1.51
	3	15.89	1.08	14.57	2.21	10.76	1.01	11.43	1.11
	4	23.25	1.06	24.93	1.91	17.84	1.30	23.72	1.42
	5	16.93	1.70	24.52	1.89	19.13	1.22	31.13	1.38
	6	7.04	1.53	12.74	1.01	13.11	1.01	17.15	1.48
VARIÂNCIA	1	4.20	1.53	8.47	1.59	4.94	1.12	6.33	1.82
	2	10.38	1.49	6.22	1.27	7.78	1.38	6.60	1.89
	3	13.79	1.27	1.67	1.13	8.99	1.35	5.75	1.58
	4	12.88	1.31	4.50	1.60	7.46	1.38	0.49	1.33
	5	7.32	1.08	9.61	1.81	4.15	1.35	2.76	1.29
	6	1.71	1.37	10.29	1.90	0.03	1.15	3.08	1.28
COEFICIENTE DE VARIÂÇÃO	1	3.97	1.58	7.74	1.33	7.33	1.19	8.36	1.52
	2	6.08	1.35	3.24	1.05	4.38	1.13	5.01	1.56
	3	5.76	1.04	2.24	1.47	2.33	1.10	1.48	1.30
	4	3.66	1.12	8.91	2.05	1.03	1.06	4.98	1.15
	5	0.50	1.59	14.20	2.22	4.36	1.23	9.28	1.04
	6	1.61	1.66	12.49	1.91	7.12	1.34	7.98	1.04

STUDENT CRÍTICO = 1.96 FISHER CRÍTICO = 1.0

TABELA (IV.4) - Pesquisa sobre a existência de Periodicidade nos parâmetros das amostras - Teste de STUDENT e FISHER

Observando-se a variação do valor de P, ainda na tabela (IV.5), nota-se um crescimento relativamente pequeno com o acréscimo de harmônicos, porém, esse fato não é suficiente para se desprezar a significância desses harmônicos tendo em vista que o somatório da variância explicada pelos mesmos poderá representar uma porcentagem grande da variação dos parâmetros. Surge, então, o problema da sub ou super estimação de harmônicos que poderá afetar o prosseguimento da análise das séries.

b) Método do Periodograma Acumulado

Analisando a figura (IV.10) podemos chegar aos resultados apresentados na tabela (IV.6), de acordo com o descrito em 3.3.2.

PARÂMETRO	ESTATÍSTICA	CAMPINAS	ECOLOGIA AGRÍCOLA
TEMPERATURA	MÉDIA	1	1
	VARIÂNCIA	1	1
RADIAÇÃO	MÉDIA	1	1
	VARIÂNCIA	1	1

TABELA IV.6 - HARMÔNICOS SIGNIFICANTES - MÉTODO PERIODOGRAMO ACUMULADO

Esses resultados também se mostraram insuficientes para a decisão a tomar quanto ao número de harmônios significantes, apenas servindo como um teste de apoio para a decisão final.

Nº DE HARMÔNIO	C A M P I N A S								E C O L O G I A A G R Í C O L A							
	R A D I A Ç Ã O				T E M P E R A T U R A				R A D I A Ç Ã O				T E M P E R A T U R A			
	Média		Variância		Média		Variância		Média		Variância		Média		Variância	
	H	P	H	P	H	P	H	P	H	P	H	P	H	P	H	P
1	1	0.6941	1	0.4187	1	0.8340	1	0.2860	1	0.6498	1	0.2180	1	0.8903	25	0.1226
2	2	0.7286	13	0.4623	2	0.8972	3	0.3200	3	0.6637	23	0.2496	2	0.9097	6	0.1707
3	4	0.7497	98	0.4814	20	0.9002	6	0.3500	7	0.6768	12	0.2759	19	0.9169	17	0.2150
4	19	0.7640	67	0.4969	8	0.9046	26	0.3737	8	0.6874	46	0.3013	10	0.9229	39	0.2527
5	23	0.7761	2	0.5117	19	0.9086	2	0.3918	44	0.6975	4	0.3262	8	0.9280	26	0.2858
6	10	0.7852	10	0.5260			42	0.4097	14	0.7060	14	0.3497			13	0.3080
7	3	0.7929	23	0.5367			35	0.4266	36	0.7137	34	0.3721			19	0.3287
8	18	0.8005	14	0.5468			54	0.4434	40	0.7214	19	0.3919			5	0.3485
9	55	0.8076	163	0.5561			94	0.4592	11	0.7286	18	0.4107			62	0.3678
10			117	0.5650			33	0.4729	2	0.7352	36	0.4273			126	0.3860
11			30	0.5746			16	0.4864	63	0.7417	7	0.4438			1	0.4039
12			119	0.5832			69	0.4996	16	0.7481	35	0.4588			20	0.4210
13			24	0.5915			53	0.5112	50	0.7541	16	0.4738			95	0.4367
14			25	0.5998			77	0.5226	17	0.7600	27	0.4874			89	0.4522
15			70	0.6073			38	0.5336	55	0.7651	87	0.5003			102	0.4655
16			106	0.6145			20	0.5447	60	0.7702	73	0.5112			37	0.4786
			
			
			
	P _{min} =0.1994		P _{min} =0.1410		P _{min} =0.1994		P _{min} =0.1410		P _{min} =0.1901		P _{min} =0.1344		P _{min} =0.1901		P _{min} =0.1344	
	P _{max} =0.806		P _{max} =0.8590		P _{max} =0.8086		P _{max} =0.8590		P _{max} =0.8099		P _{max} =0.8656		P _{max} =0.8099		P _{max} =0.8656	

TABELA IV.5 - TESTE DE HARMÔNICOS SIGNIFICANTES - YEVJWICH -

H - ordem sequencial da harmônica P - Somatório do parâmetro ΔP_j (item 3.3.2)

c) Método dos intervalos de confiança

Pelo fato das séries diárias exigirem a computação de 182 harmônicas, implicando em muito tempo de máquina, não foi possível empregar fielmente o método dos intervalos de confiança. Então, procedeu-se de maneira inversa, ou seja: a componente periódica foi calculada, agregando-se sucessivamente, os harmônicos do mais significativo para o menos significativo. Os harmônicos significativos, ficariam determinados quando somente 5% ou menos dos 365 valores da componente periódica estivessem fora dos limites estabelecidos.

A tabela (IV.7) apresenta os harmônicos significativos inferidos pelo método dos intervalos de confiança, para o nível de 95%.

PARÂMETRO	ESTATÍSTICA	CAMPINAS		ECOLOGIA AGRÍCOLA	
		HS	VFL	HS	VFL
TEMPERATURA	Média	6	18	5	21
	Variância	4	17	5	17
RADIAÇÃO	Média	2	17	7	19
	Variância	10	26	9	30

TABELA IV.7 - Harmônicas significativas - Método dos intervalos de confiança (HS - harmônicas significativas, VFL número de valores fora dos limites)

O intervalo de confiança fixado em 95% permite que 18 valores caiam fora dos limites. Examinando a tabela, (IV.7) verifica-se que somente em quatro casos se obtém o resultado esperado. Nos outros casos, esses resultados foram os que se apre

sentaram menos inconvenientes, tendo em vista que agregando-se mais harmônicos o número de valores fora dos limites diminuía muito pouco, não justificando o acréscimo de harmônicos.

Nenhum dos 3 testes aplicados forneceu resultado preciso e então em presença da incerteza dos mesmos foi necessário lançar mão de um critério subjetivo, tomando os resultados apresentados pelos três testes como apoio e também valendo-se do ajuste gráfico entre a componente periódica computada com números de harmônicos significantes diferentes, e as médias e desvios padrões amostrais.

Foi decidido usar as seguintes harmônicas significantes com base nos resultados obtidos:

Radiação Solar Global:

Estação Centro Experimental de Campinas:

- 5 harmônicas na média, e
- 7 harmônicas na variância

Estação Ecologia Agrícola:

- 2 harmônicas na média, e
- 7 harmônicas na variância

Temperatura:

Estação Centro Experimental de Campinas:

- 6 harmônicas na média, e
- 4 harmônicas na variância

Estação Ecologia Agrícola:

- 5 harmônicas na média,
- 5 harmônicas na variância

Nas figuras (IV.2 a IV.9) estão plotadas as componentes periódicas na média e desvio padrão para as quatro séries. Na tabela (IV.8) são apresentados os coeficientes de Fourier calculados conforme as equações (III.14 e III.15) que entraram no cálculo das componentes periódicas.

4.2.4 - Análise da componente estocástica dependente e seleção do modelo autoregressivo

Determinada a componente periódica na média e desvio padrão, esta última por meio da variância, obteve-se a componente estocástica com o uso do modelo descrito pela equação (III.8) para todas as séries em estudo.

A média e a variância da componente estocástica nas quatro séries foram analisadas quanto a estacionaridade pelo método da análise de harmônicas não sendo detectada frequência dominante.

Para determinação da ordem do modelo de dependência foi necessário, primeiramente, investigar a possível existência de periodicidade nos coeficientes de autocorrelação das séries estimados através da equação (III.30). A investigação da periodicidade nos 1º e 2º coeficientes de correlação serial foi efetuada dividindo a amostra em duas subamostras de acordo com o item (3.3.3.3) aplicando-se em seguida os testes de igualdade de parâmetros de Student e Fisher (item 3.3.1) às mesmas. A tabela (IV.10) apresenta os resultados desses testes para as quatro séries. Conforme observado nesta tabela somente a série de radiação solar global da Estação Centro Experimental de Campi

NÚMERO DE HARMÔ- NICOS		COEFICIENTES DE FOURIER											
		CAMPINAS						ECOLOGIA AGRÍCOLA					
		MÉDIA			VARIÂNCIA			MÉDIA			VARIÂNCIA		
		OH	A	B	OH	A	B	OH	A	B	OH	A	B
RADIÇÃO SOLAR GLOBAL	1	1	71.60	-1.81	1	5024.64	-2259.86	1	79.55	21.66	1	6057.95	-3633.26
	2	2	-14.45	6.81	13	-373.82	-1738.48	3	11.71	-2.90	23	-1055.29	-2475.73
	3	4	-8.43	9.22	98	75.66	1173.48				12	-1058.91	-2216.40
	4	19	10.05	2.56	67	544.07	911.38				46	-2329.84	-615.84
	5	23	-9.47	-0.10	2	-312.27	-988.44				4	1960.72	-1364.56
	6				10	-698.67	-741.39				14	-315.35	-2297.49
	7				23	838.30	279.99				34	-2150.62	712.59
TEMPERATURA	1	1	-2.72	1.14	1	0.31	-1.15	1	2.63	0.58	1	-1.01	-1.94
	2	2	-0.16	0.40	17	-0.73	-0.13	2	-0.59	0.44	3	0.51	0.56
	3	19	-0.16	0.21	3	0.69	0.17	20	0.15	0.13	6	0.66	-0.27
	4	10	-0.14	0.20	23	-0.66	0.06	8	-0.04	-0.19	26	-0.62	-0.11
	5	8	0.06	-0.21				19	-0.09	0.16	2	-0.55	0.04
	6	5	-0.13	0.15									

TABELA IV.8 - COEFICIENTES DE FOURIER E HARMÔNICOS SIGNIFICANTES. (OH)-ordem do harmônico

nas não passou no teste. Porém, como não foi possível remover a não homogeneidade desta série, a periodicidade nos coeficientes $r_{k,\zeta}$ desta série será atribuída a este fato, e para fins deste trabalho, esses coeficientes serão considerados não periódicos. As figuras (IV.11 a IV.16) mostram a variação desses coeficientes. O passo seguinte consistiu em determinar a ordem do modelo autoregressivo e estimar os coeficientes de autocorrelação adotados para a computação dos coeficientes autoregressivos do modelo.

POSTO	PARÂMETRO	1º Coefic. de Correl.		2º Coef. de Cor	
		t	F	t	F
CAMPINAS	TEMPERATURA	1.57	0.73	1.54	1.15
	RADIAÇÃO	0.001	1.58	1.10	1.30
ECOLOGIA AGRÍCOLA	TEMPERATURA	1.49	0.36	1.00	1.12
	RADIAÇÃO	0.074	1.10	0.16	1.20
t crítico = 1.96		F crítico = 1.20			

TABELA (IV.10) - TESTE DE STUDENT E FISHER NO 1º e 2º COEFICIENTE DE CORRELAÇÃO

A primeira tentativa para selecionar a ordem do modelo foi feita empregando o teste simplificado de Yevjevich (equações III.44 a 46). Os valores de D_1 até 3a. ordem assim como a ordem do modelo selecionado são mostrados na tabela (IV.11). Os resíduos obtidos com a aplicação dos modelos selecionados por este teste foram analisados quanto a independência pelo teste de Siddiqui (equação III.51). Nenhuma das séries residuais passou no teste de independência, com aproximadamente,

40% dos pontos do correlograma caindo fora dos limites de Siddiqui, nas quatro séries.

A inspeção visual dos correlogramas da componente estocástica dependente (Figura IV.15) sugere as mesmas ordens para os modelos autoregressivos encontradas com o teste de Yevjevich.

Visando obter uma componente estocástica independente a seleção dos modelos foi então realizada com o seguinte procedimento: foram computadas as componentes presumidamente independentes através dos modelos autoregressivos (equação III.33) até 6a. ordem para as quatro séries considerando-se as duas alternativas de uso dos coeficientes de autocorrelação serial (item 3.3.3.3). Em seguida foram obtidos os correlogramas das séries residuais e aplicado o teste de Siddiqui, que fornece os limites de tolerância para os valores do correlograma. Com esse procedimento, melhores resultados eram obtidos quando se utilizava os coeficientes autoregressivos $\bar{r}_{1,\zeta}$, $\bar{r}_{2,\zeta}$ embora a correspondente diferença entre estes e os r_1 , r_2 , fosse pequena.

POSTO	PARÂMETRO	D1	D2	D3	MODELO AUTOREGRESSIVO
CAMPINAS	RADIAÇÃO SOLAR	0.1989	0.2033	0.2102	1 ^a ordem
	TEMPERATURA	0.4585	0.4604	0.4607	1 ^a ordem
ECOLOGIA AGRÍCOLA	RADIAÇÃO SOLAR	0.3380	0.3529	0.3586	2 ^a ordem
	TEMPERATURA	0.5034	0.5134	0.5163	1 ^a ordem

TABELA IV.11 - DETERMINAÇÃO DA ORDEM DO MODELO AUTOREGRESSIVO
CRITÉRIO YEVJEVICH

A figura (IV.16) mostra os correlogramas das componentes estocásticas independentes para as quatro séries, obtidos de acordo com a equação (III.30), com um defasamento de 100 dias e os limites obtidos para os mesmos pelo teste de Siddiqui. A tabela (IV.13) apresenta os coeficientes autoregressivos empregados no modelo, obtidos conforme equação (III.37), e os desvios padrões obtidos com as expressões (III.41 a III.43).

As estatísticas da componente estocástica independente são dadas na tabela (IV.14).

	PARAMETRO	COEFICIENTES AUTOREGRESSIVOS						DESVIO PADRÃO
		α_1	α_2	α_3	α_4	α_5	α_6	
CAMPINAS	RADIAÇÃO SOLAR	0.3898	0.0339	0.0633	0.0384	0.0249	0.0625	0.78
	TEMPERATURA	0.7173	-0.0743	-0.0058	0.0449			0.54
ECOLOGIA AGRÍCOLA	RADIAÇÃO SOLAR	0.4430	0.0792	0.0190	0.0720	0.0730	0.0871	0.62
	TEMPERATURA	0.8186	-0.1984	0.0543	0.0282			0.48

TABELA IV.13 - COEFICIENTES AUTOREGRESSIVOS DOS MODELOS SELECIONADOS

A tabela (IV.12) apresenta os valores encontrados para r_1 , r_2 , r_3 , que são os três primeiros coeficientes de autocorrelação da série $\epsilon_{p,\tau}$ integral e $\bar{r}_{1,\tau}$, $\bar{r}_{2,\tau}$ e $\bar{r}_{3,\tau}$ que são médias dos coeficientes de autocorrelação serial na posição τ .

COEFICIENTE DE AUTOCORRELAÇÃO	CAMPINAS		ECOLOGIA AGRÍCOLA	
	RADIAÇÃO SOLAR	TEMPERATURA	RADIAÇÃO SOLAR	TEMPERATURA
$\bar{r}_{1,\tau}$	0.445	0.677	0.581	0.709
r_1	0.430	0.677	0.567	0.701
$\bar{r}_{2,\tau}$	0.263	0.427	0.437	0.433
r_2	0.234	0.411	0.409	0.408
$\bar{r}_{3,\tau}$	0.218	0.280	0.364	0.288
r_3	0.186	0.261	0.339	0.265

TABELA IV.12 - COEFICIENTES DE AUTOCORRELAÇÃO DA SÉRIE $\epsilon_{p,\tau}$

Os modelos autoregressivos selecionados empregaram então os coeficientes autoregressivos $\bar{r}_{1,\zeta}$, $\bar{r}_{2,\zeta}$... e a ordem dos mesmos foi determinada quando o maior número de pontos do correlograma do resíduo permaneceu dentro dos limites obtidos pela expressão (III.51). Os modelos autoregressivos selecionados foram de 6ª ordem para as duas séries de radiação solar e 4ª ordem para as duas séries de temperatura. A seleção de ordem foi feita utilizando o critério do teste de independência de Siddiqui, como foi explicado anteriormente.

POSTO	PARÂMETRO	MÉDIA	VARIÂNCIA	COEFICIENTE DE ASSIMETRIA	COEFICIENTE DE CURTOSE
CAMPI <u>N</u> AS	RADIAÇÃO	- 0.0010	1.0233	- 0.7554	0.4950
	TEMPERA <u>T</u> URA	0.0001	0.9937	- 0.6927	4.3280
ECOLO <u>G</u> IA AGRÍCO <u>L</u> IA	RADIAÇÃO	0.0005	1.0342	- 0.3248	3.7181
	TEMPERA <u>T</u> URA	- 0.0005	1.0143	- 0.3934	4.2283

TABELA IV.14 - ESTATÍSTICAS DA COMPONENTE INDEPENDENTE

Para verificação da estacionaridade dos modelos autoregressivos selecionados foram construídas as equações características (equações IV.3 a IV.6) e para suas resoluções empregou-se a subrotina POLTR do Statistical Scientific Package-IBM implantada na biblioteca do Núcleo de Computação Eletrônica da UFRJ. Para que o modelo seja estacionário é necessário que o módulo das raízes seja maior que 1 (um), e isso se verifica conforme pode ser observado na tabela (IV.16).

As equações características de radiação solar global e de temperatura para o posto Estação Centro Experimental de Campinas, são respectivamente:

$$1 - 0.3898 B - 0.0339 B^2 - 0.0633 B^3 - 0.0384 B^4 - 0.0249 B^5 - 0.0625 B^6 = 0 \quad (\text{IV.3})$$

$$1 - 0.7173 B + 0.0743 B^2 + 0.0058 B^3 - 0.0449 B^4 = 0 \quad (\text{IV.4})$$

As equações características de radiação solar global e de temperatura para o posto Ecologia Agrícola, são respectivamente:

$$1 - 0.4430 B - 0.0792 B^2 - 0.0190 B^3 - 0.0720 B^4 - 0.0730 B^5 - 0.0871 B^6 = 0 \quad (\text{IV.5})$$

$$1 - 0.08186 B + 0.1984 B^2 - 0.0543 B^3 - 0.0282 B^4 = 0 \quad (\text{IV.6})$$

4.2.5 - Ajuste de uma distribuição de probabilidade a componente estocástica independente

As quatro distribuições descritas no item (3.3.4) foram usadas no ajuste das distribuições de frequências da componente estocástica independente.

Para a realização do teste de qui-quadrado (item 3.3.5) foram usados os intervalos de classe de igual probabilidade, sendo a probabilidade de cada intervalo de classe igual a

$$P_j = \frac{1}{k}, \text{ com } j = 1, 2, \dots, k \quad (\text{IV.7})$$

onde k é o intervalo de classe selecionado. O intervalo de classe foi selecionado empregando-se a equação (III.67) que resultou em 28 classes. Esse número de intervalos de classe foi usado no teste de qui-quadrado para todas as distribuições descritas. Nenhuma passou no teste de aderência. Os valores encontrados para a estatística χ^2 com 28 classes ficaram no intervalo de 100 a 500 e os valores críticos para o nível de significância de 95% estão entre 20 e 30, de acordo com tabelas apresentadas em Neville, 1964. Nova tentativa foi feita empregando

	RAÍZES DO POLINÔMIO						
	B_1	B_2	B_3	B_4	B_5	B_6	
C A M P I N A S	PARTE REAL	1.220	-1.783	0.843	0.843	-0.759	-0.759
	PARTE IMA- GINÁRIA	0.	0.	-1.359	1.359	-1.518	1.518
	MÓDULO	1.220	1.783	1.598	1.598	1.697	1.697
	PARTE REAL	1.384	0.876	0.876	-3.007	-	-
	PARTE IMA- GINÁRIA	0.	-2.141	2.141	0.	-	-
	MÓDULO	1.384	2.313	2.313	3.007	-	-
	PARTE REAL	1.106	-1.691	0.728	-0.728	-0.855	-0.855
	PARTE IMA- GINÁRIA	0.	0.	-1.239	1.239	-1.497	1.497
	MÓDULO	1.106	1.106	1.437	1.437	1.724	1.724
ECOLOGIA AGRÍCOLA	PARTE REAL	1.376	0.786	0.786	-4.872	-	-
	PARTE IMA- GINÁRIA	0.	-2.162	2.162	0.	-	-
	MÓDULO	1.376	2.299	2.299	4.872		

TABELA IV.15 - RAÍZES DA EQUAÇÃO CARACTERÍSTICA DOS
MODELOS AUTOREGRESSIVOS SELECIONADOS

20 intervalos de classe e igualmente as estatísticas χ^2 superaram em muito os valores tabelados. Persistindo na tentativa, e empregando somente a função densidade de probabilidade gama "double branch", foram usados 32 e 40 intervalos de classe. Os valores encontrados para χ^2 ficaram entre 100 e 300 o que está muito além dos χ^2 críticos. Portanto, não foi possível ajustar uma distribuição de frequência a componente estocástica independente. As figuras (IV.21 e IV.22) mostram a distribuição de frequência da componente estocástica independente para as quatro séries, onde pode ser visto que as mesmas possuem picos altos e pontiagudos no centro, e caudas longas. Nas figuras (IV.17 a IV.20) estão as distribuições de frequências acumuladas e as funções de distribuição normal e "gamma double-branch" acumuladas ajustadas.

4.2.6 - Discussão dos resultados

Quando se formula um modelo matemático para investigar um fenômeno, pretende-se que este ajude a descrever, da melhor maneira possível o fenômeno em estudo. Entretanto, a escolha do modelo adequado está condicionada a julgamento críticos que naturalmente são variáveis. Esses julgamentos críticos geralmente podem ser apoiados em testes estatísticos ou em experiências anteriores.

O bom resultado do modelo depende basicamente da qualidade e quantidade dos dados pois estes influem significativamente na seleção e estimação de seus parâmetros. Para se obter um modelo adequado é recomendável dispor de uma série longa

de observações, homogênea e sem falhas. Períodos curtos de observações podem levar, por exemplo, a uma possível falsa identificação de tendências e saltos devido às variações aleatórias que está sujeito o fenômeno. Assim, em nosso caso, como só se dispunha de curto período de dados (10 e 11 anos) não foi possível identificar conclusivamente uma possível tendência na série de radiação solar global de Campinas. Caso a mesma realmente exista, os resultados da análise da estrutura desta série apresentados neste trabalho devem ser corrigidos. Por outro lado, a forma do preenchimento das falhas utilizada aparentemente não teve muita influencia nos resultados visto a similaridade destes e o grau de dificuldade encontrado na determinação dos modelos e harmônicas significantes para a série de temperatura da Estação Centro Experimental de Campinas onde não houve nenhuma falha.

Ficou evidenciada durante a apresentação dos resultados a dificuldade encontrada na determinação do número de harmônicas significantes. Por outro lado, cabe ressaltar que em séries de radiação solar global e temperatura fica muito difícil se tentar justificar fisicamente mais de uma harmônica, tendo em vista que estas variáveis estão diretamente relacionadas ao período anual de translação da terra.

Neste trabalho, como as experiências anteriores são poucas, decidiu-se considerar o teste de Siddiqui como o instrumento adequado para medir o bom ajuste dos modelos autoregressivos, embora suas premissas não sejam satisfeitas completamente e apresente limites bastante estreitos, principalmente no ca

so das séries diárias de radiação solar global.

Embora tenham sido selecionados modelos autoregressivos de 6a. e 4a. ordem, os valores dos coeficientes desses modelos indicam que o dia imediatamente anterior é o de maior influencia. A temperatura se mostrou mais dependente do dia anterior do que a radiação solar global, com os valores de r_1 para temperatura nos dois postos sendo 0.68 e 0.70 e para radiação solar global nos dois postos sendo 0.43 e 0.57. Isto pode ser explicado pelo fato de que a temperatura tem um tempo de resposta maior do que a radiação solar global quando estão atuando fatores aperiódicos, como é o caso da ocorrência de nebulosidade e opacidade da atmosfera.

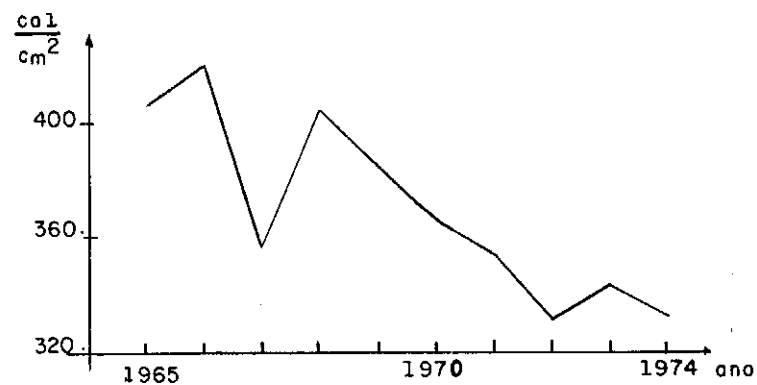
Do exposto acima, admite-se que a ordem dos modelos autoregressivos selecionados, principalmente para a radiação solar global, podem não ser suficientemente adequadas para estas variáveis, porém a certeza só poderia vir com uma disponibilidade maior de dados e testes mais adequados.

A ordem dos modelos autoregressivos e a determinação do número de harmônicos significantes tem importancia fundamental na estimação da série residual. A forma das distribuições das frequencias das componentes estocásticas independentes apresentaram, como já foi visto, picos altos e pontiagudos e caudas longas. Isto poderia ser produzido por um ajuste inadequado da componente periódica conforme conclusões do trabalho de Bullard, 1976. Porém, não se pode garantir que isto seja em decorrência de harmônicas super ou sub estimadas, tendo em vista que (como

já foi mencionado no Capítulo II) foram encontradas as mesmas formas de distribuição de frequência para componentes estocásticas independentes de séries diárias homogêneas com longo período de observação.

Nenhuma das funções de distribuição de probabilidades empregadas para ajuste da distribuição de frequências passou no teste de qui-quadrado para os níveis de significância de 0.05 e 0.01. A ordem de grandeza dos qui-quadrados calculados é a mesma para todas as funções de distribuição de probabilidade usadas.

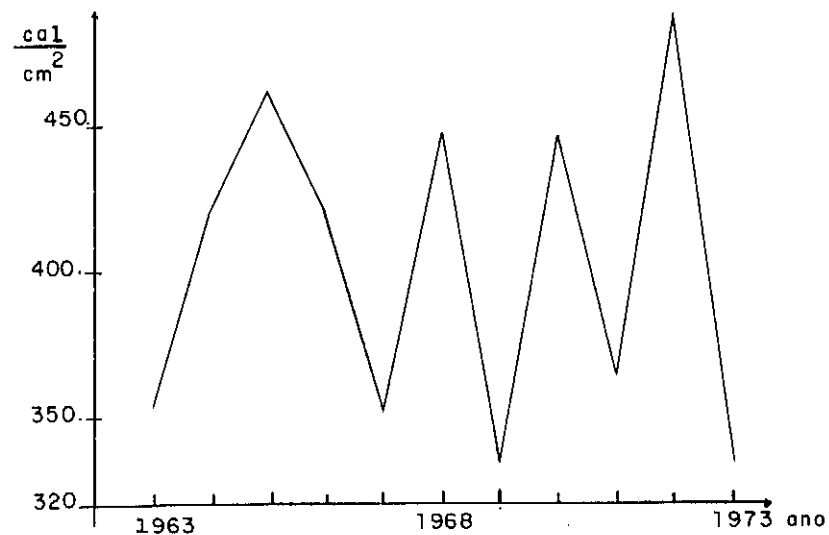
Apesar das dificuldades indicadas nesta discussão, considera-se que os modelos apresentados com seus respectivos parâmetros podem representar satisfatoriamente as séries analisadas.



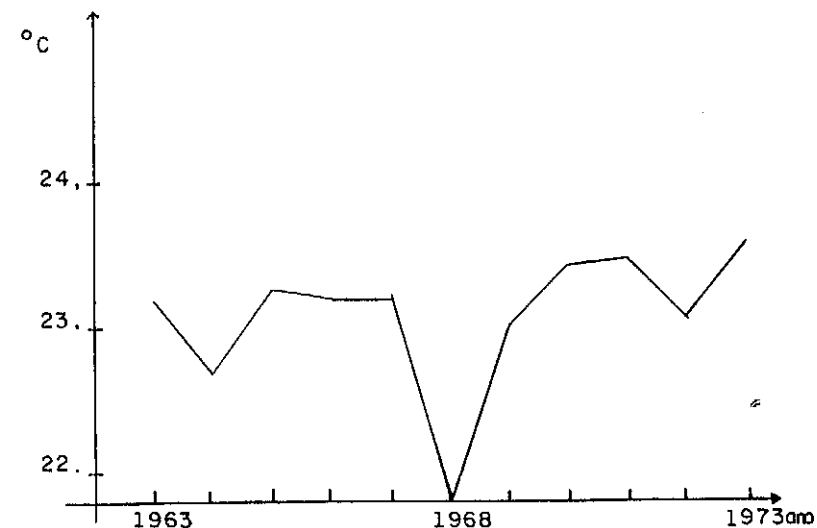
(a) Rad. Solar - Est. Campinas



(b) Temperatura - Est. Campinas



(c) Rad. Solar - Est. Ecologia Agrícola



(d) Temperatura - Est. Ecologia Agrícola

FIGURA IV.1 - SÉRIES ANUAIS DE RADIAÇÃO SOLAR GLOBAL E TEMPERATURA MÉDIA DO AR.

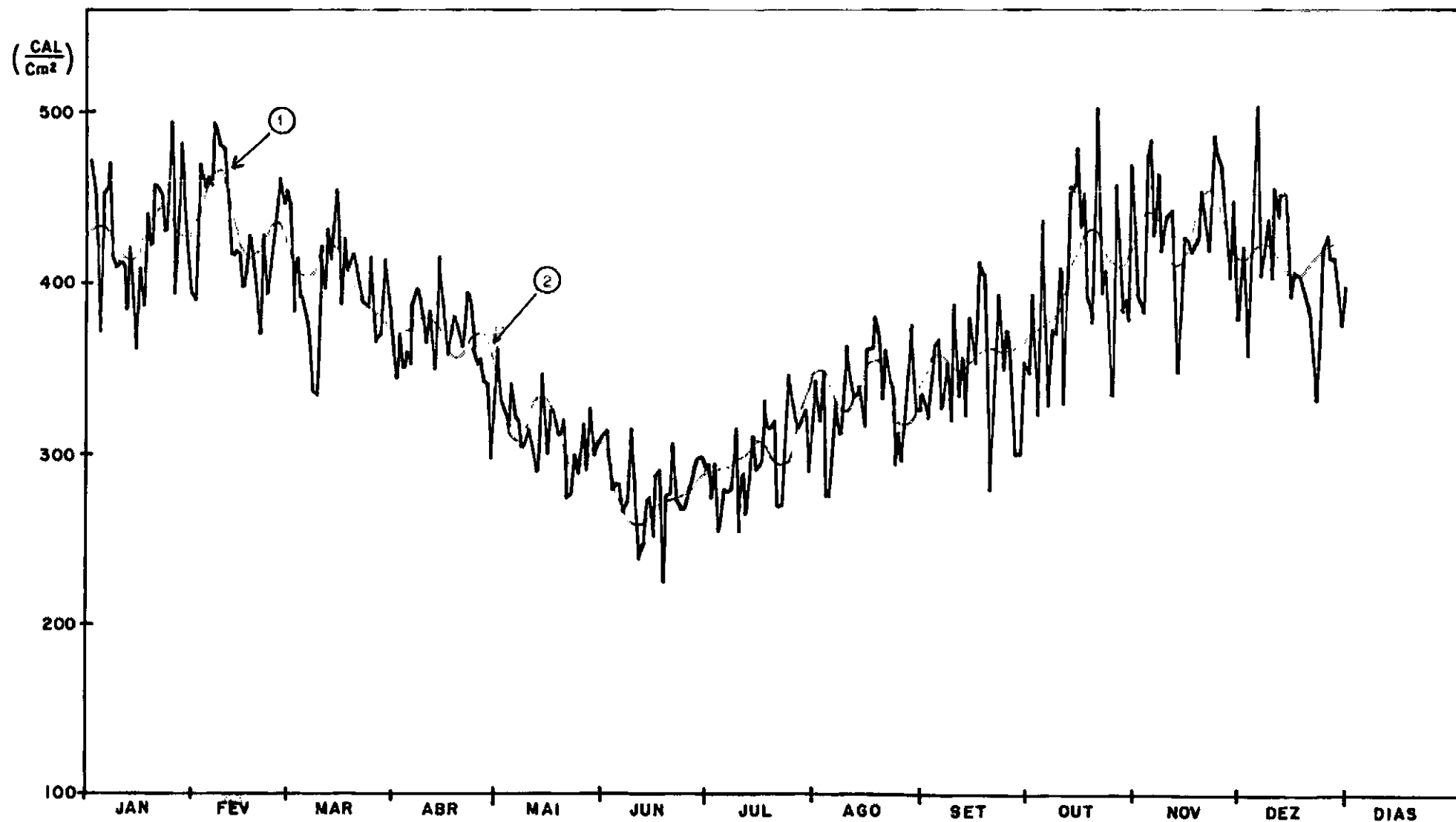


FIGURA (IV.2) – ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS. (1) MÉDIA DIÁRIA E (2) COMPONENTE PERIÓDICA AJUSTADA COM 5 HARMÔNICAS DA RADIAÇÃO SOLAR GLOBAL DIÁRIA, PERÍODO 1965 a 1974.

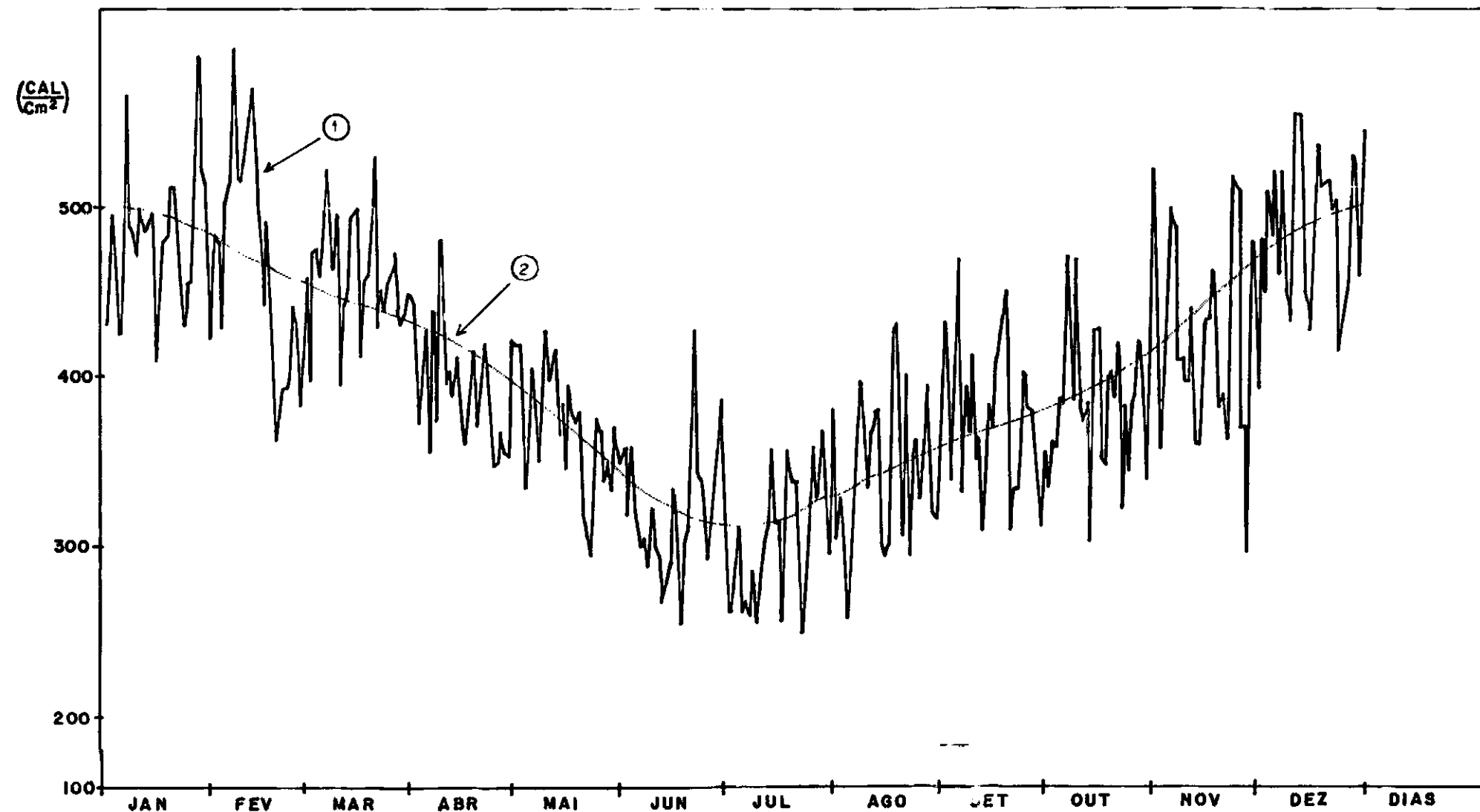


FIGURA (IV.3) - ESTAÇÃO ECOLOGIA AGRÍCOLA. (1) MÉDIA DIÁRIA E (2) COMPONENTE PERIÓDICA AJUSTADA COM 2 HARMÔNICAS DA RADIAÇÃO SOLAR GLOBAL DIÁRIA, PERÍODO 1963 a 1973.

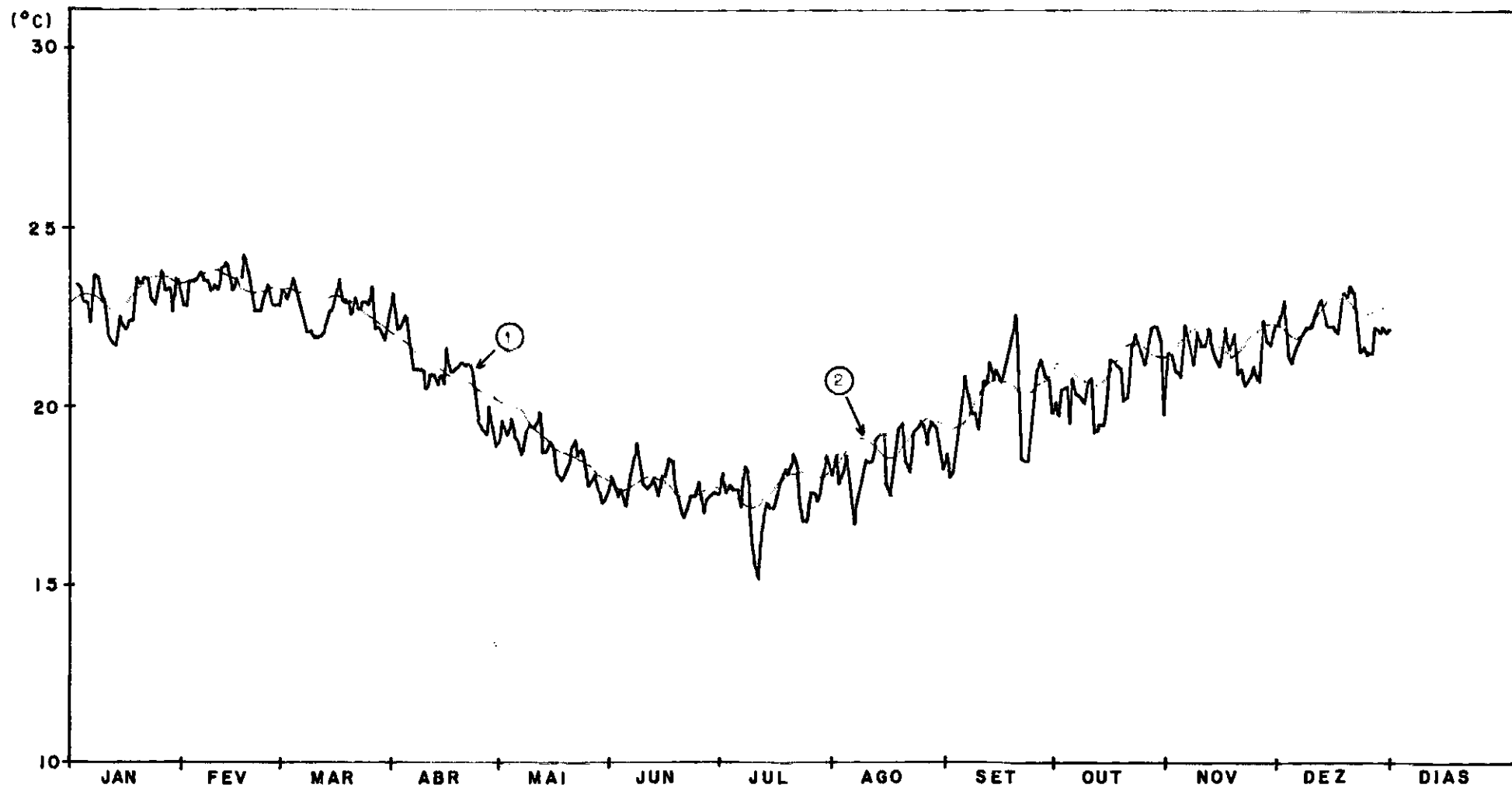


FIGURA (IV.4) - ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS. (1) MÉDIA DIÁRIA E (2) COMPONENTE PERIÓDICA AJUSTADA COM 6 HARMÔNICAS DA TEMPERATURA MÉDIA DIÁRIA, PERÍODO 1965 a 1974.

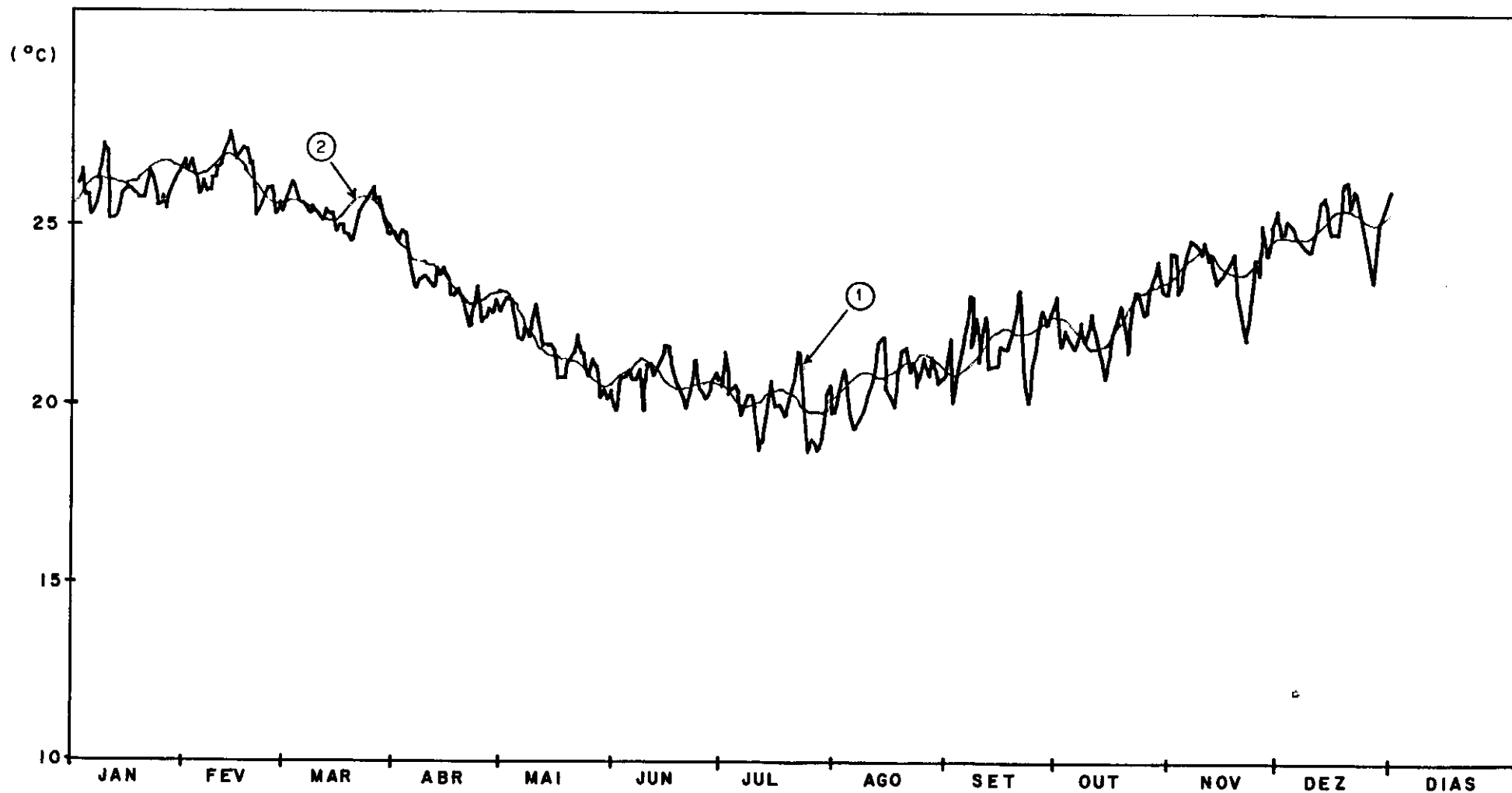


FIGURA (IV.5) — ESTACÃO ECOLOGIA AGRÍCOLA. (1) MÉDIA DIÁRIA E (2) COMPONENTE PERIÓDICA AJUSTADA COM 5 HARMÔNICAS DA TEMPERATURA MÉDIA DIÁRIA, PERÍODO 1963 a 1973.

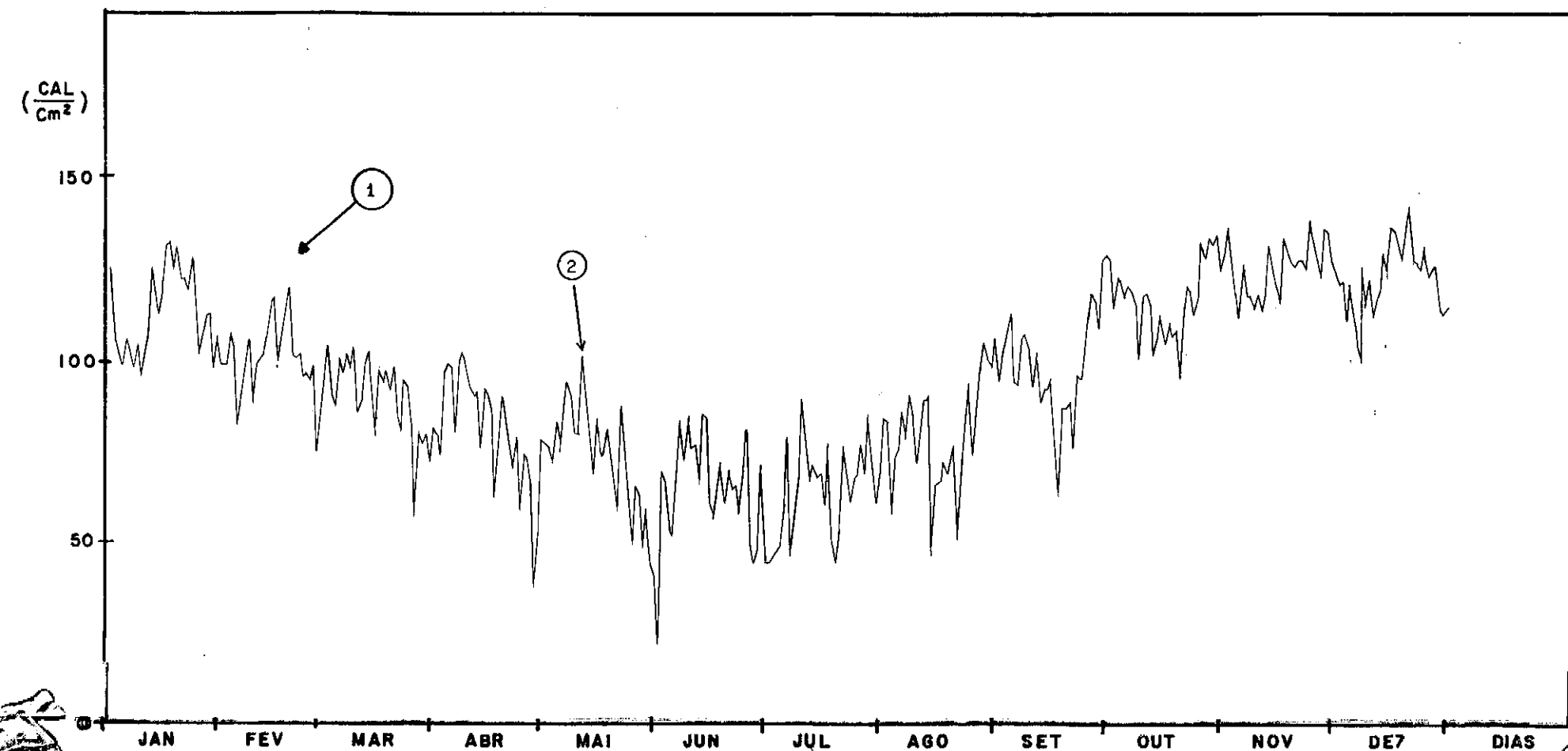


FIGURA (IV.6) - ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS. (1) DESVIO PADRÃO DIÁRIO E (2) COMPONENTE PERIÓDICA AJUSTADA COM 7 HARMÔNICAS DA RADIAÇÃO SOLAR GLOBAL DIÁRIA PERÍODO 1965 a 1974.

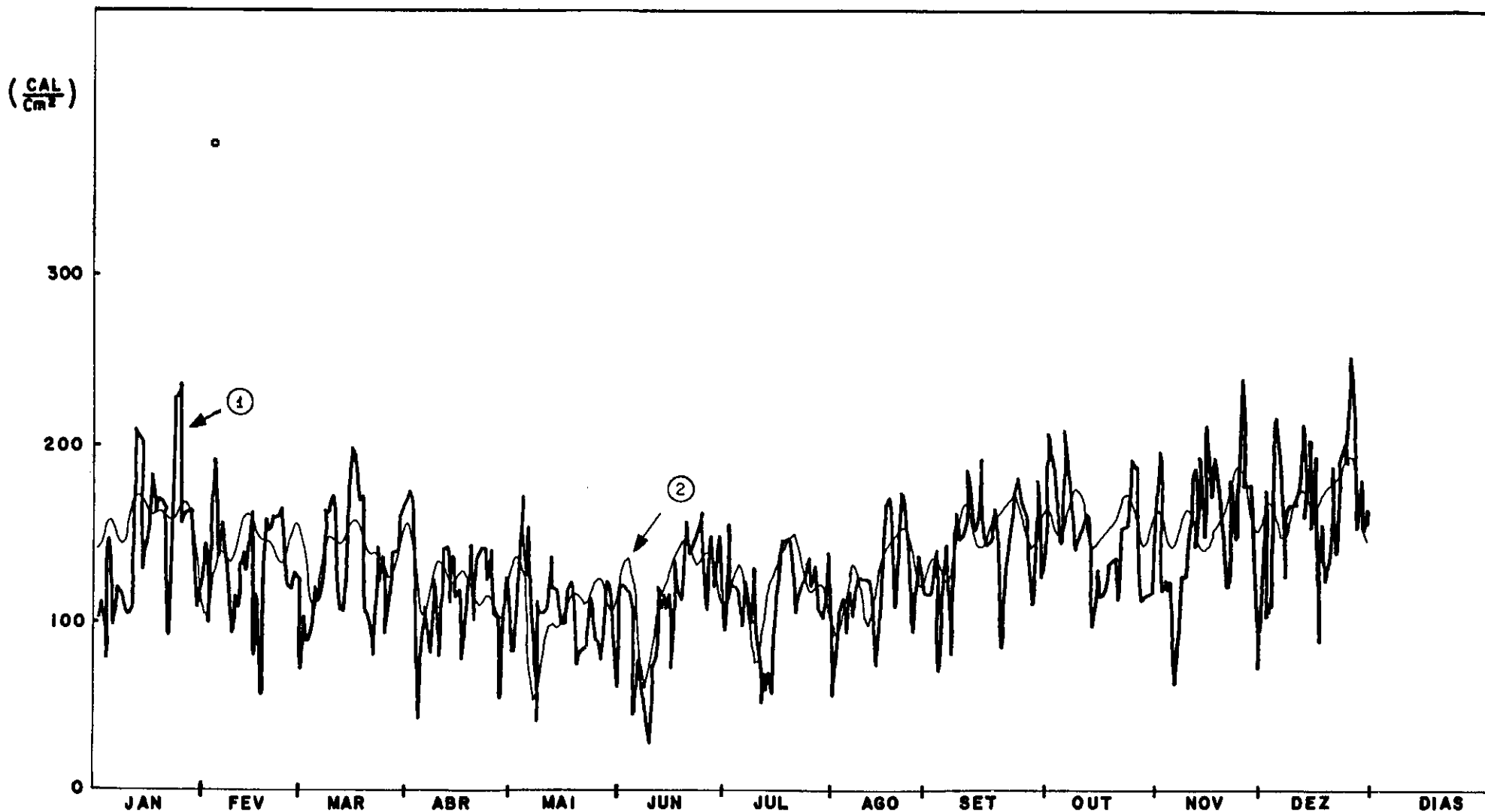


FIGURA (IV.7) - ESTAÇÃO ECOLOGIA AGRÍCOLA. (1) DESVIO PADRÃO DIÁRIO E (2) COMPONENTE PERIÓDICA AJUSTADA COM 7 HARMÔNICAS DA RADIAÇÃO SOLAR GLOBAL DIÁRIA, PERÍODO 1963 a 1973.

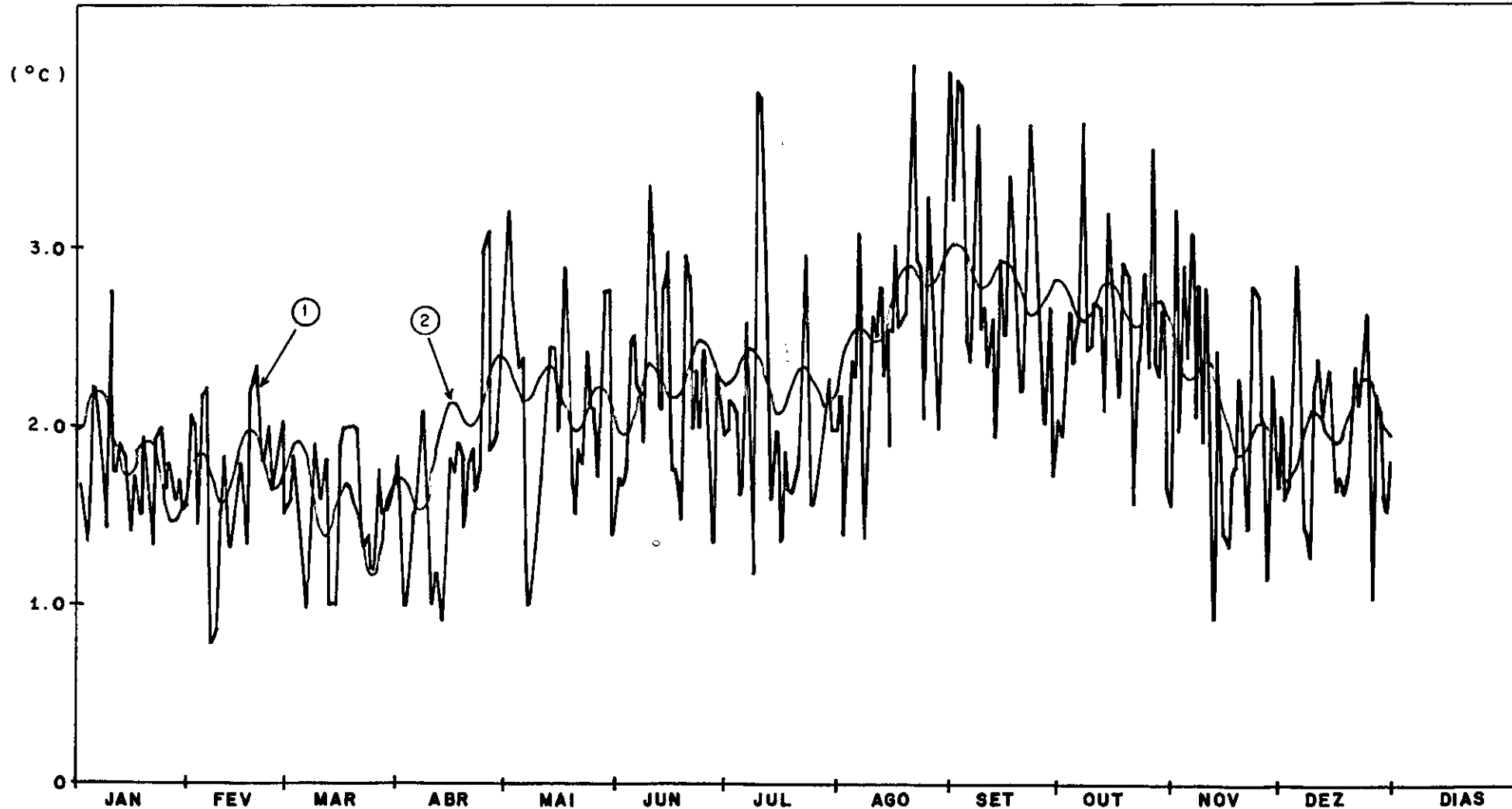


FIGURA (IV.8) - ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS. (1) DESVIO PADRÃO MÉDIO DIÁRIO E (2) COMPONENTE PERIÓDICA AJUSTADA COM 4 HARMÔNICAS DA TEMPERATURA MÉDIA DIÁRIA, PERÍODO 1965 a 1974.

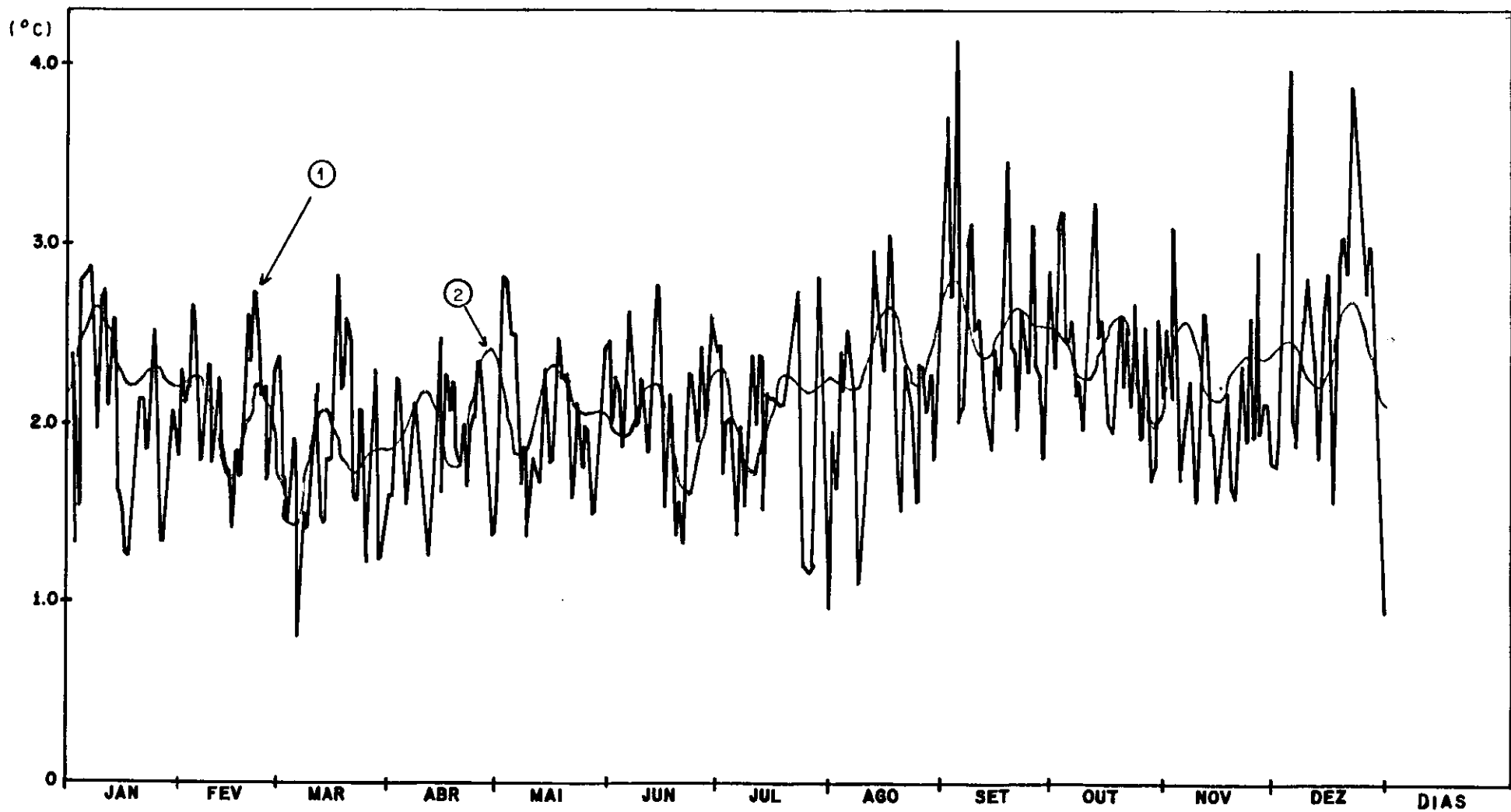


FIGURA (IV.9)—ESTAÇÃO ECOLOGIA AGRÍCOLA. (1) DESVIO PADRÃO MÉDIO DIÁRIO E (2) COMPONENTE PERIÓDICA AJUSTADA COM 5 HARMÔNICAS DA TEMPERATURA MÉDIA DIÁRIA, PERÍODO 1963 a 1974.

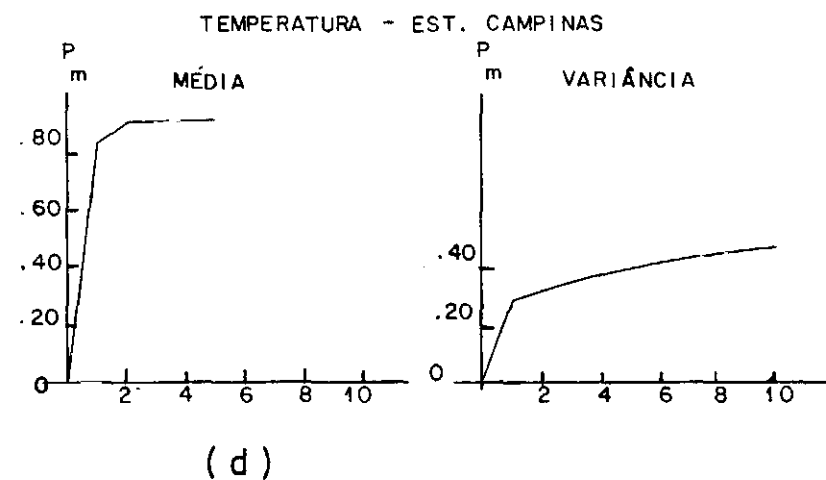
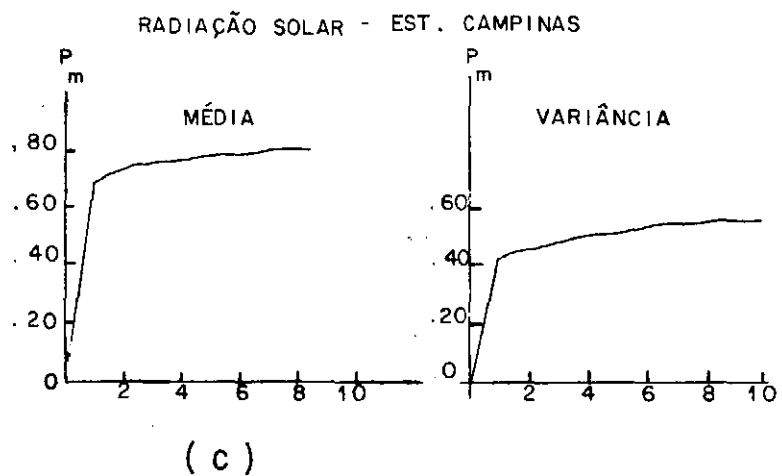
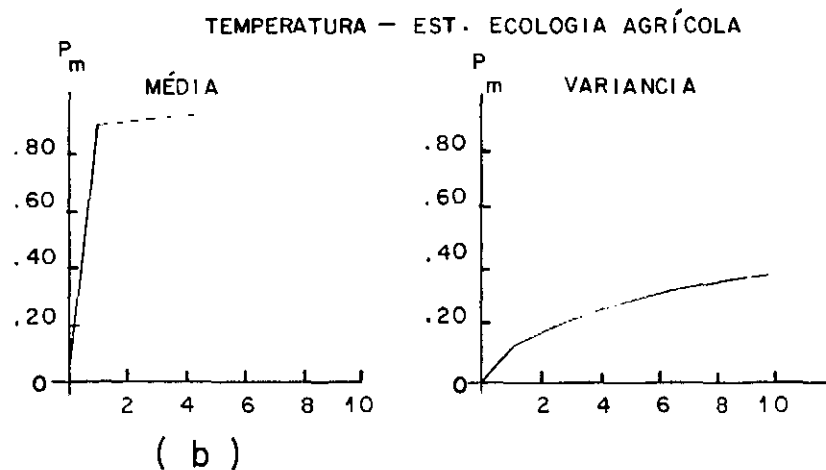
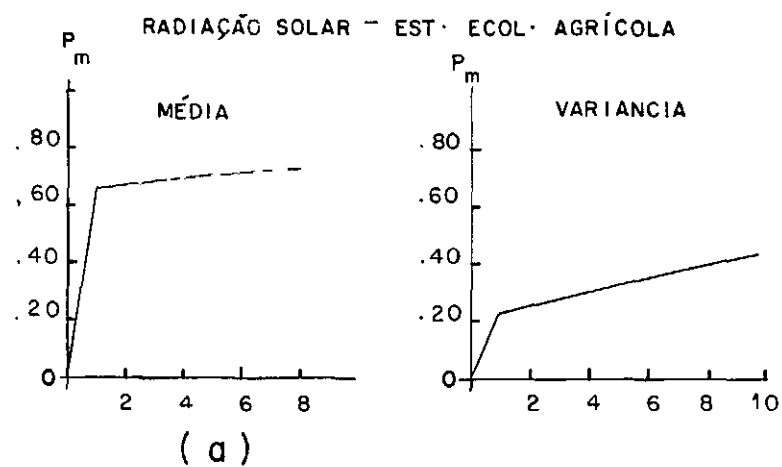


FIGURA IV.10- PERIODOGRAMA ACUMULADO DA MÉDIA E VARIÂNCIA DAS SÉRIES
RADIAÇÃO SOLAR GLOBAL E TEMPERATURA MÉDIA DO AR.

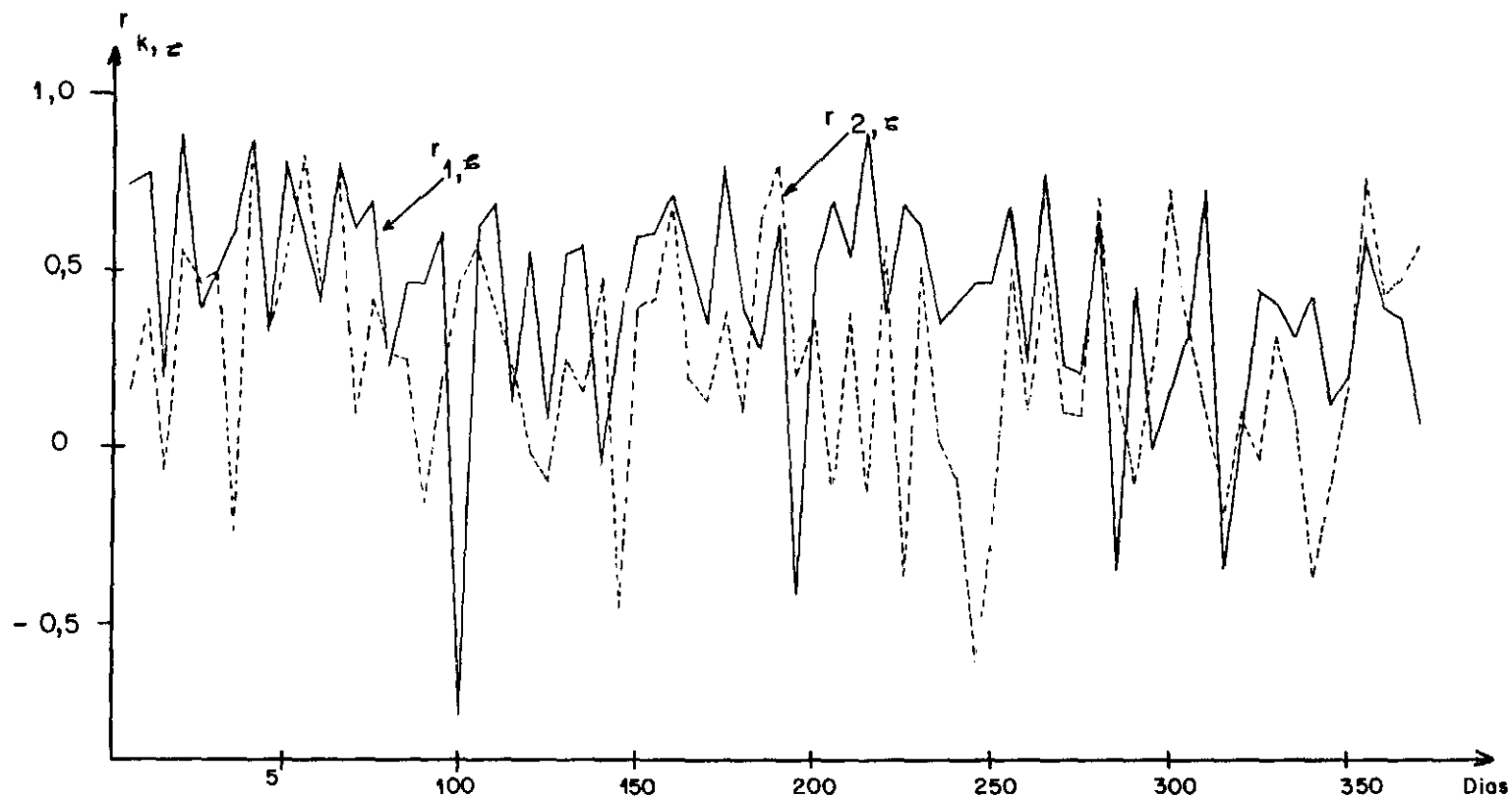


FIGURA IV.11 - 1ª E 2ª COEFICIENTES DE AUTOCORRELAÇÃO DIÁRIOS DA RADIAÇÃO SOLAR GLOBAL, ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS.

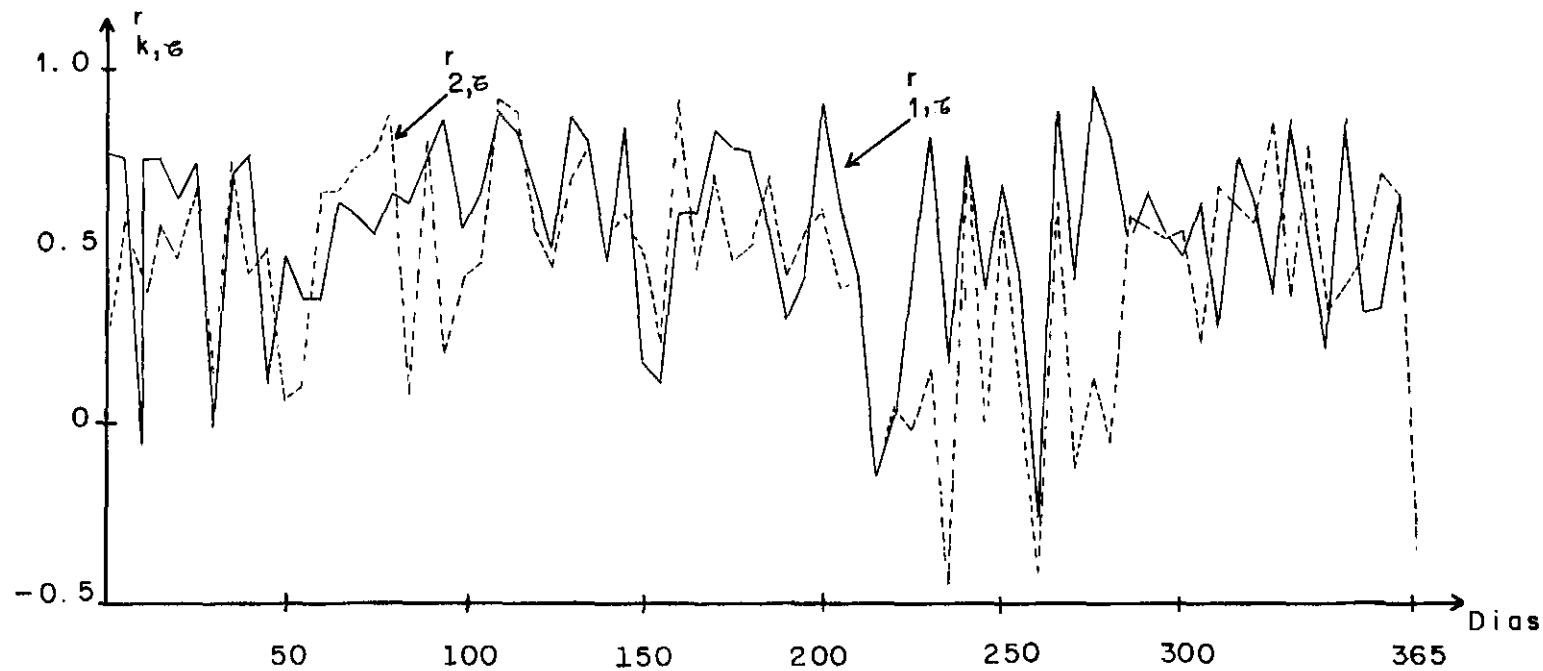


FIGURA IV.12 - 1º E 2º COEFICIENTES DE AUTOCORRELAÇÃO DIÁRIOS DA RADIAÇÃO SOLAR GLOBAL, ESTAÇÃO ECOLOGIA AGRÍCOLA

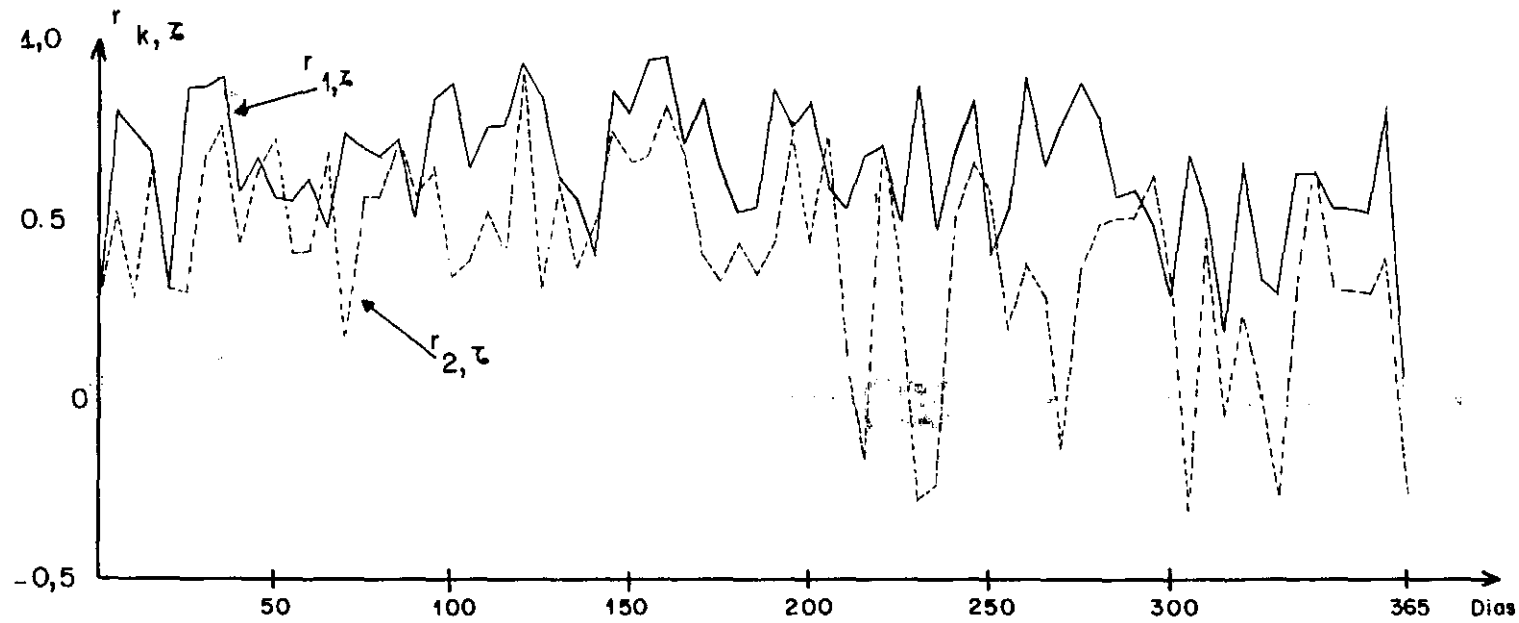


FIGURA IV.13 - 1º e 2º COEFICIENTES DE AUTOCORRELAÇÃO DIÁRIOS DA TEMPERATURA MÉDIA DIÁRIA, ESTAÇÃO CENTRO EXPERIMENTAL DE CAMPINAS.

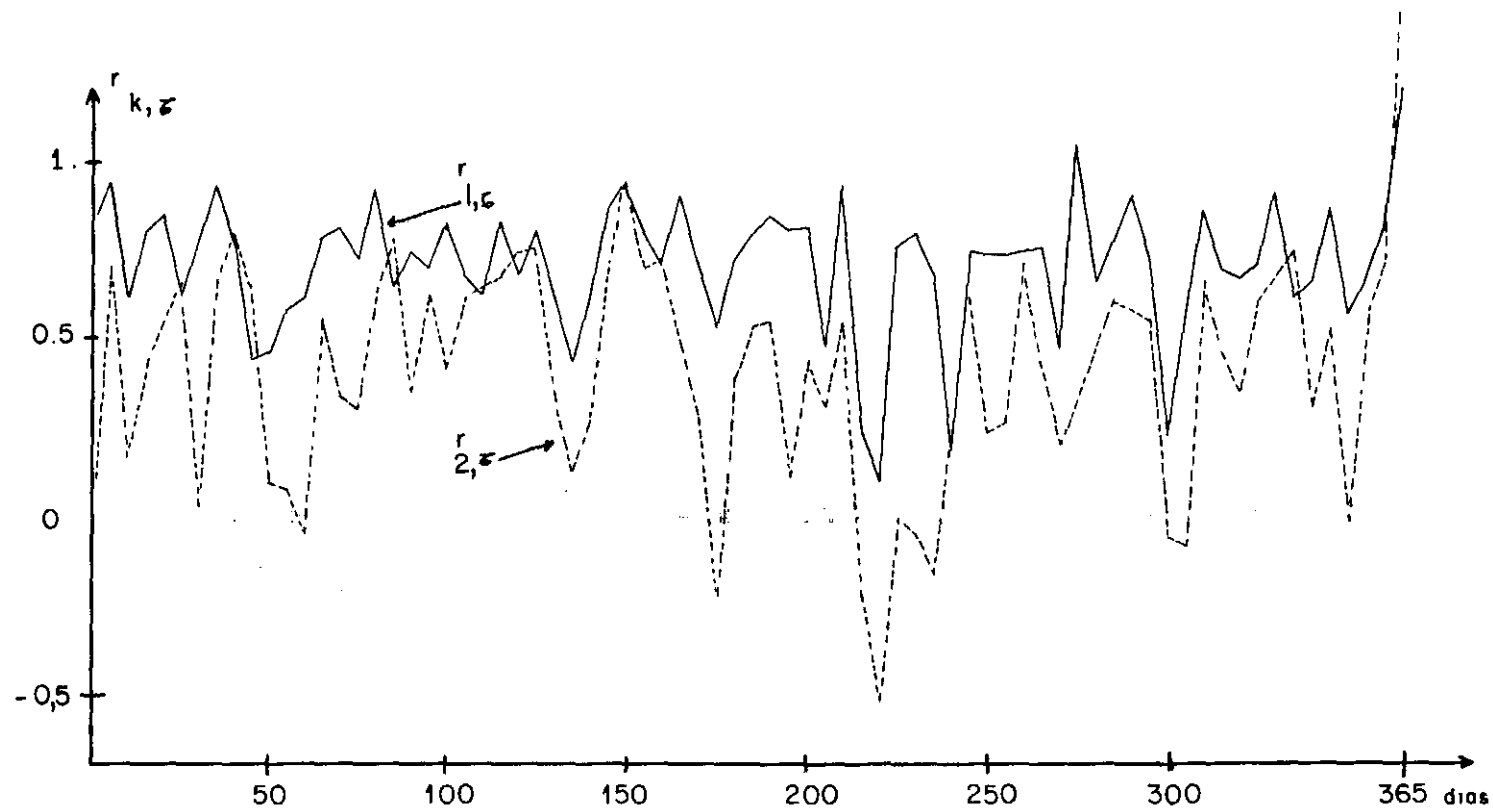


FIGURA IV.14 - 1º E 2º COEFICIENTES DE AUTOCORRELAÇÃO DIÁRIOS DA
TEMPERATURA MÉDIA DIÁRIA, ESTAÇÃO ECOLOGIA AGRÍCOLA.

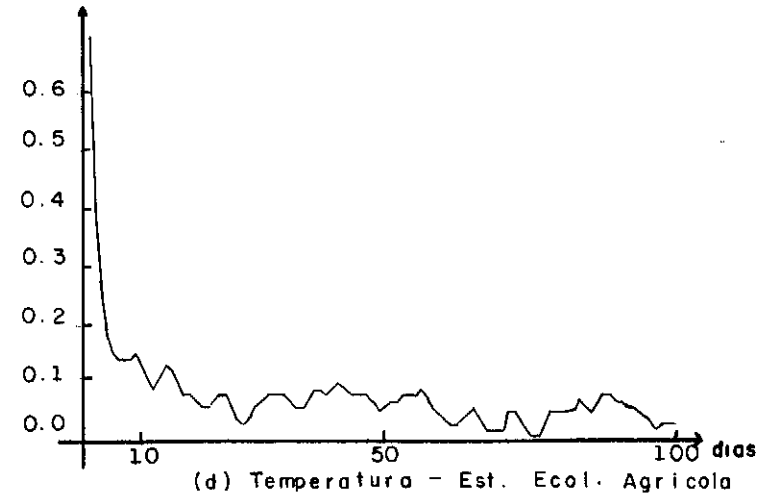
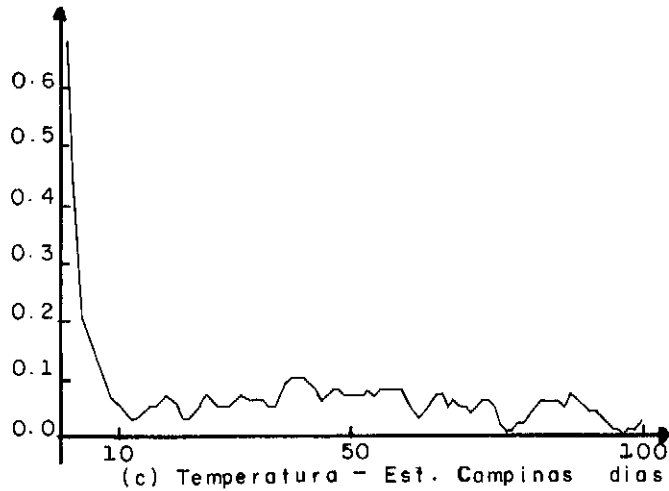
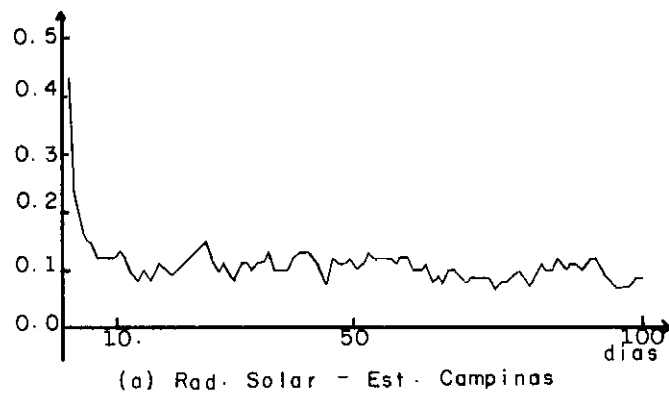


FIGURA IV.15 - CORRELOGRAMA DA COMPONENTE ESTAOCÁSTICA DEPENDENTE

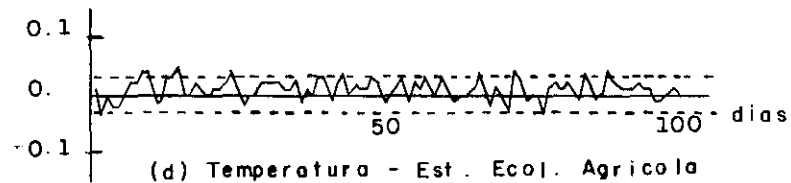
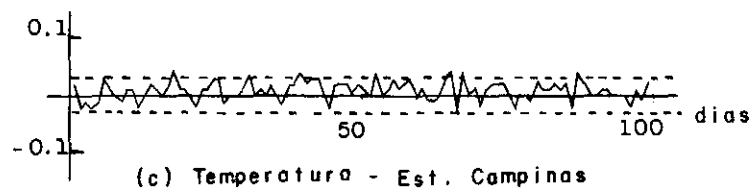
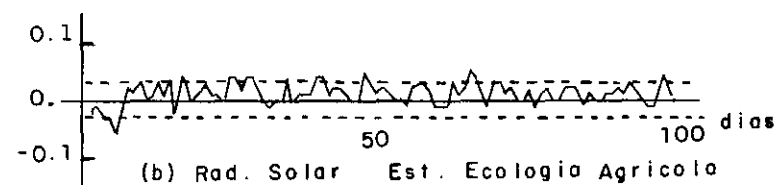
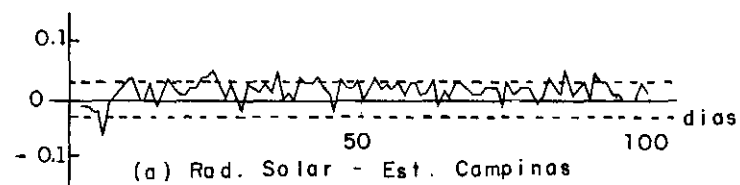
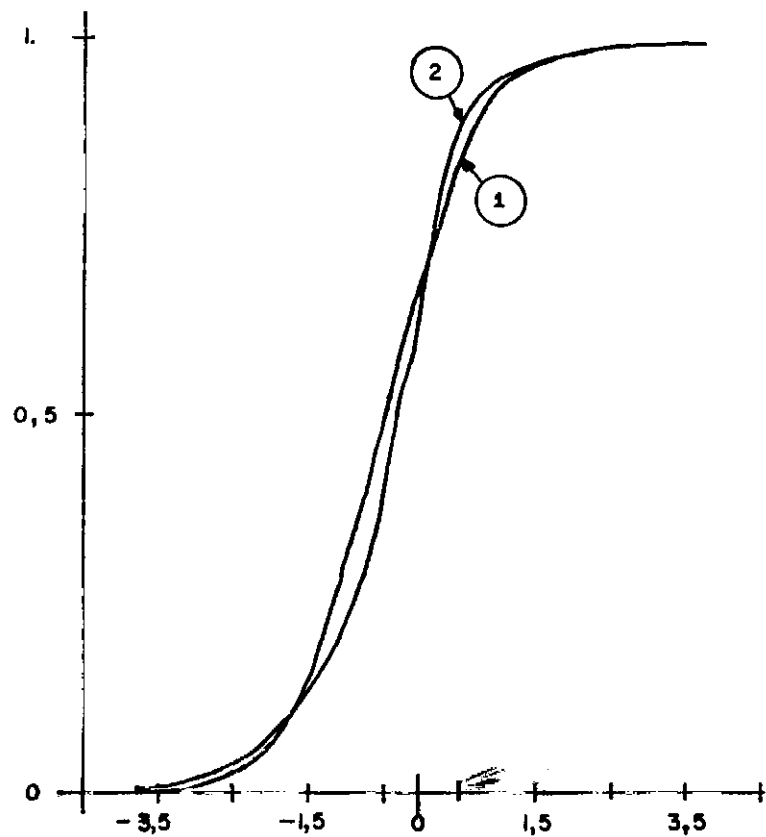
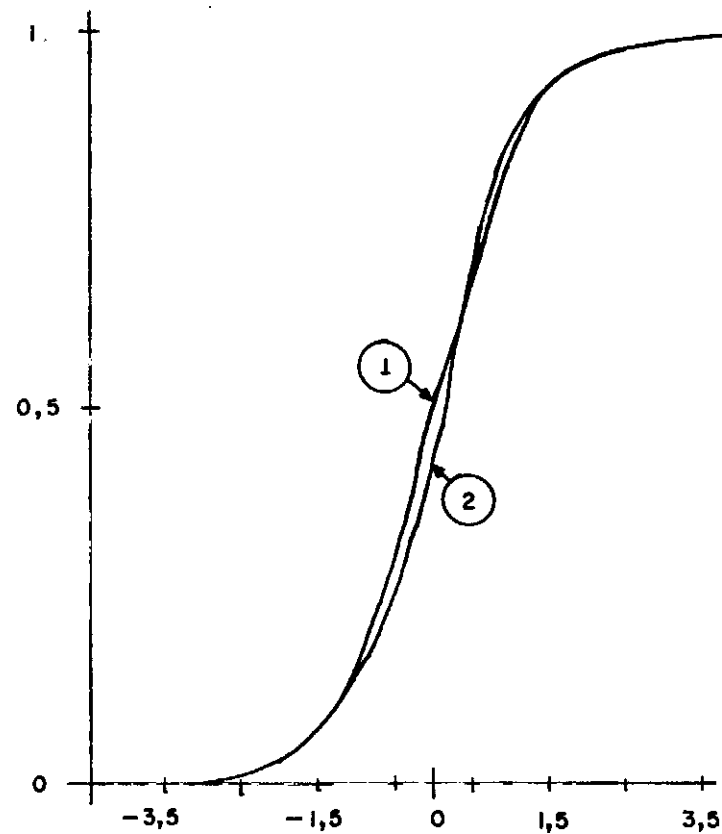


FIGURA IV.16 - CORRELOGRAMA DA COMPETENTE ESTOCÁSTICA INDEPENDENTE E SEUS RESPECTIVOS LIMITES OBTIDOS PELO TESTE DE SIDDQUI

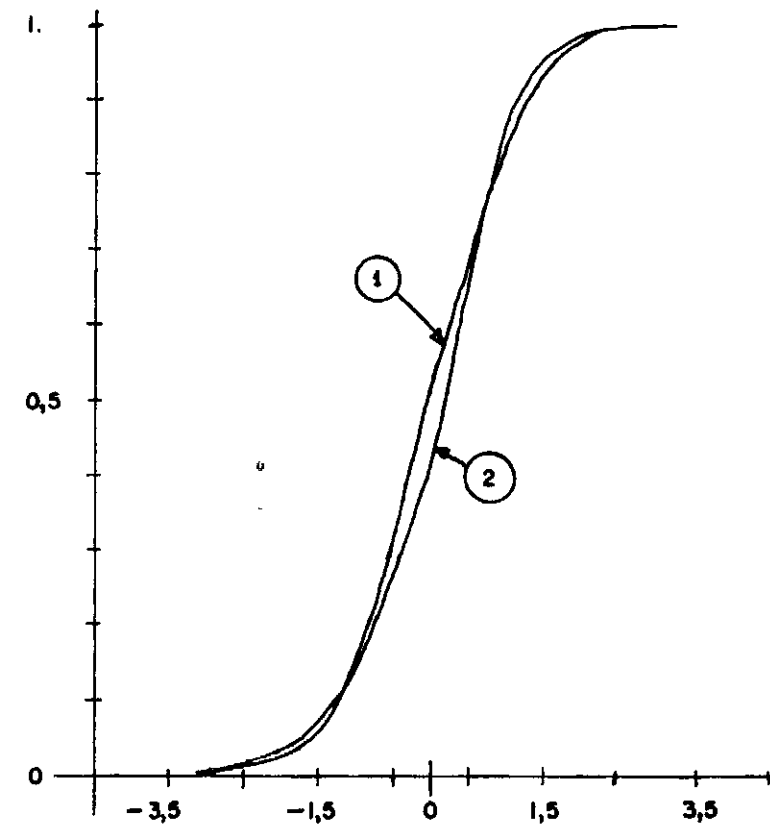


(I)

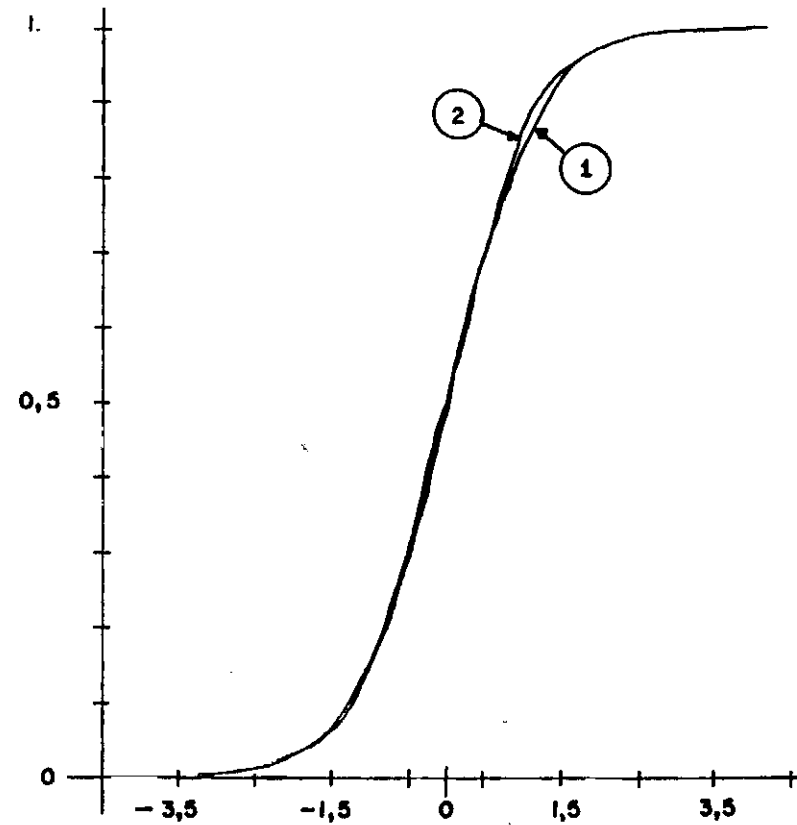


(II)

FIGURA IV.17 - FUNÇÃO DE DISTRIBUIÇÃO NORMAL ACUMULADA (I) AJUSTADA A DISTRIBUIÇÃO DE FREQUENCIAS ACUMULADA (2)
DA COMPONENTE ESTOCÁSTICA INDEPENDENTE DAS SÉRIES DIÁRIAS DE RADIAÇÃO SOLAR GLOBAL DOS POSTOS
(I) CENTRO EXPERIMENTAL DE CAMPINAS SP. E (II) ECOLOGIA AGRÍCOLA RJ.

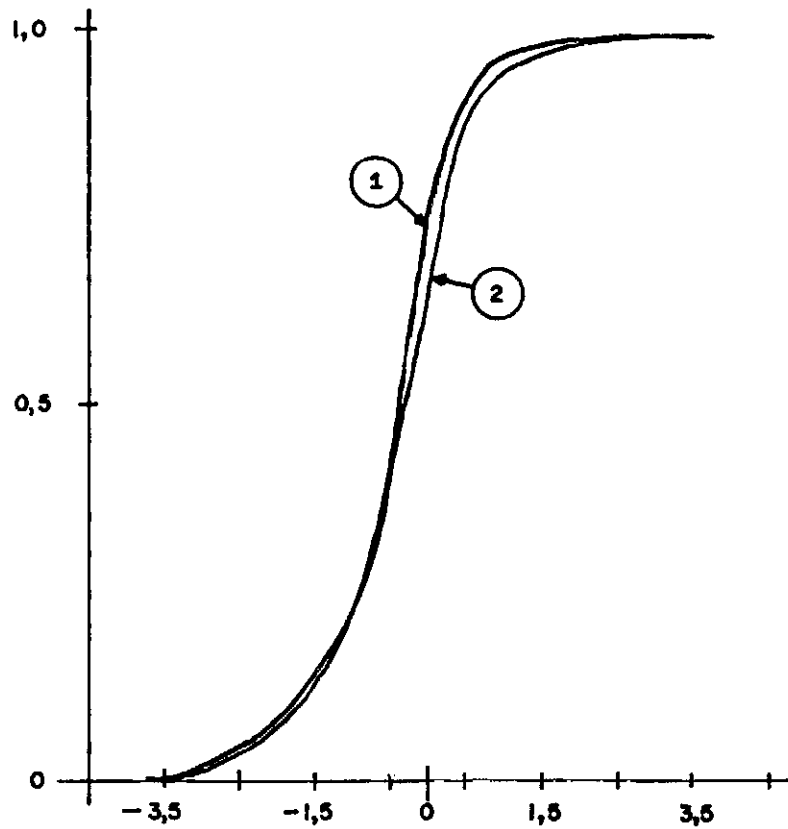


(I)

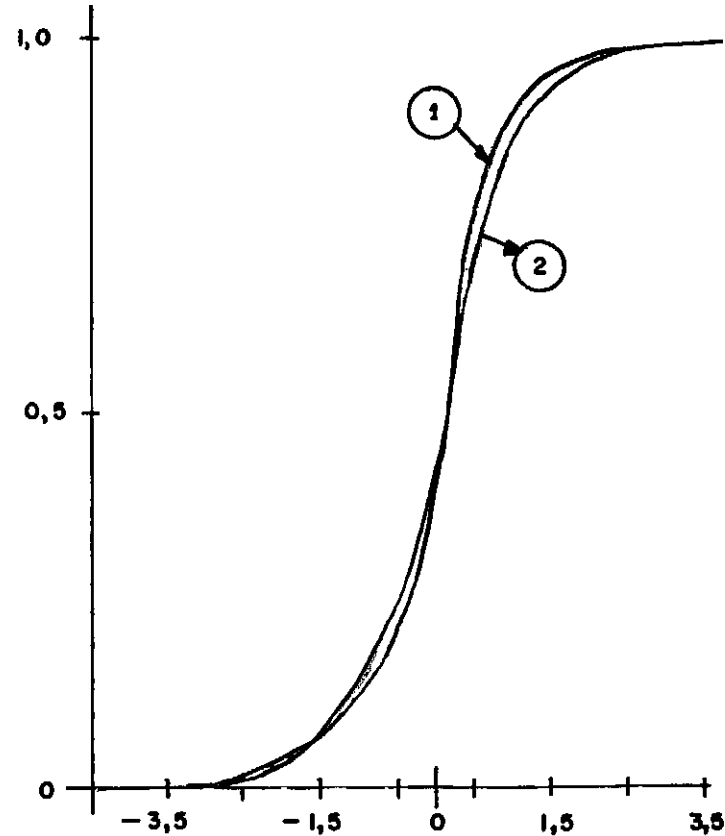


(II)

FIGURA IV.18 - FUNÇÃO DE DISTRIBUIÇÃO NORMAL ACUMULADA (1) AJUSTADA À DISTRIBUIÇÃO DE FREQUÊNCIAS ACUMULADAS (2) DA COMPONENTE ESTOCASTICA INDEPENDENTE DAS SÉRIES DE TEMPERATURA MÉDIA DIÁRIA DO POSTOS (I) CENTRO EXPERIMENTAL DE CAMPINAS-SP E (II) ECOLOGIA AGRÍCOLA-RJ.



(I)



(II)

FIGURA IV.19 - FUNÇÃO DE DISTRIBUIÇÃO "GAMA DOUBLE BRANCH" ACUMULADA (1) AJUSTADA À DISTRIBUIÇÃO DE FREQUÊNCIAS ACUMULADA (2) DA COMPONENTE ESTOCÁSTICA INDEPENDENTE DAS SÉRIES DIÁRIAS DE RADIAÇÃO SOLAR GLOBAL DOS PONTOS (1) CENTRO EXPERIMENTAL DE CAMPINAS - SP E (II) ECOLOGIA AGRÍCOLA - RJ.

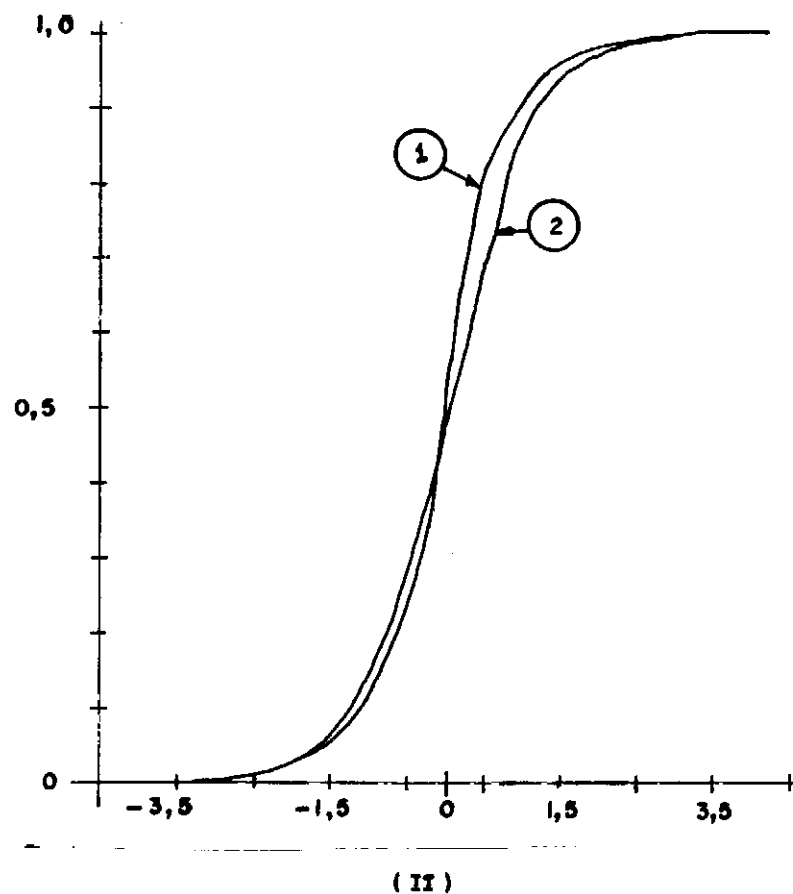
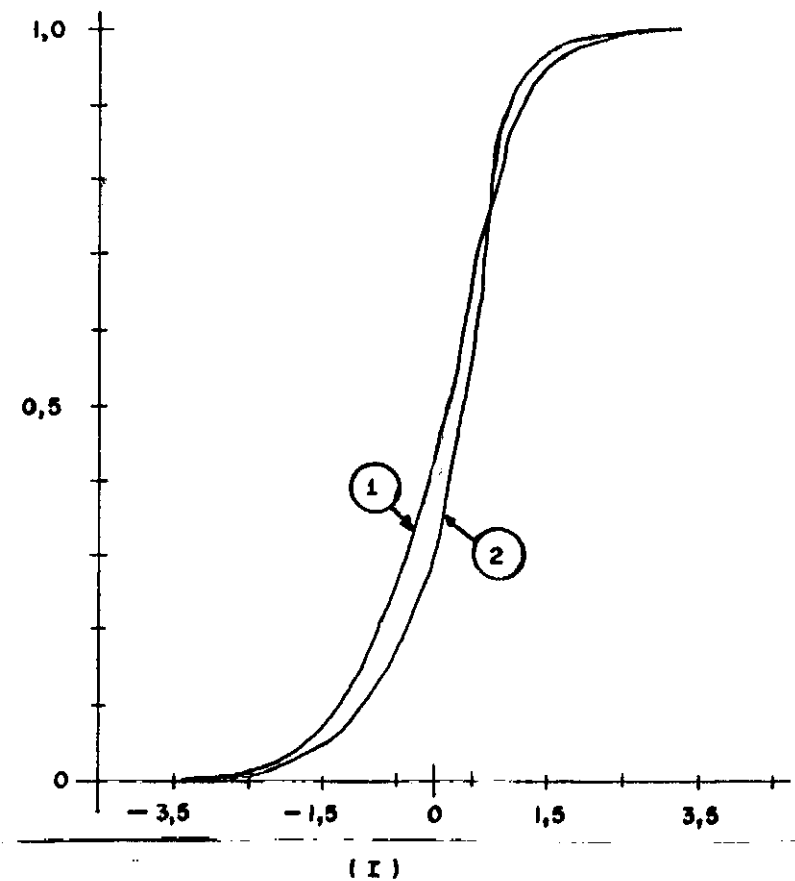


FIGURA IV.20- FUNÇÃO DE DISTRIBUIÇÃO " GAMA DOUBLE BRANCH " ACUMULADA (1) AJUSTADA À DISTRIBUIÇÃO DE FREQUÊNCIAS ACUMULADAS (2) DA COMPONENTE ESTOCÁSTICA INDEPENDENTE DAS SÉRIES DE TEMPERATURA MÉDIA DIÁRIA DOS POSTOS (I) CENTRO EXPERIMENTAL DE CAMPINAS-SP E (II) ECOLOGIA AGRÍCOLA- RJ.

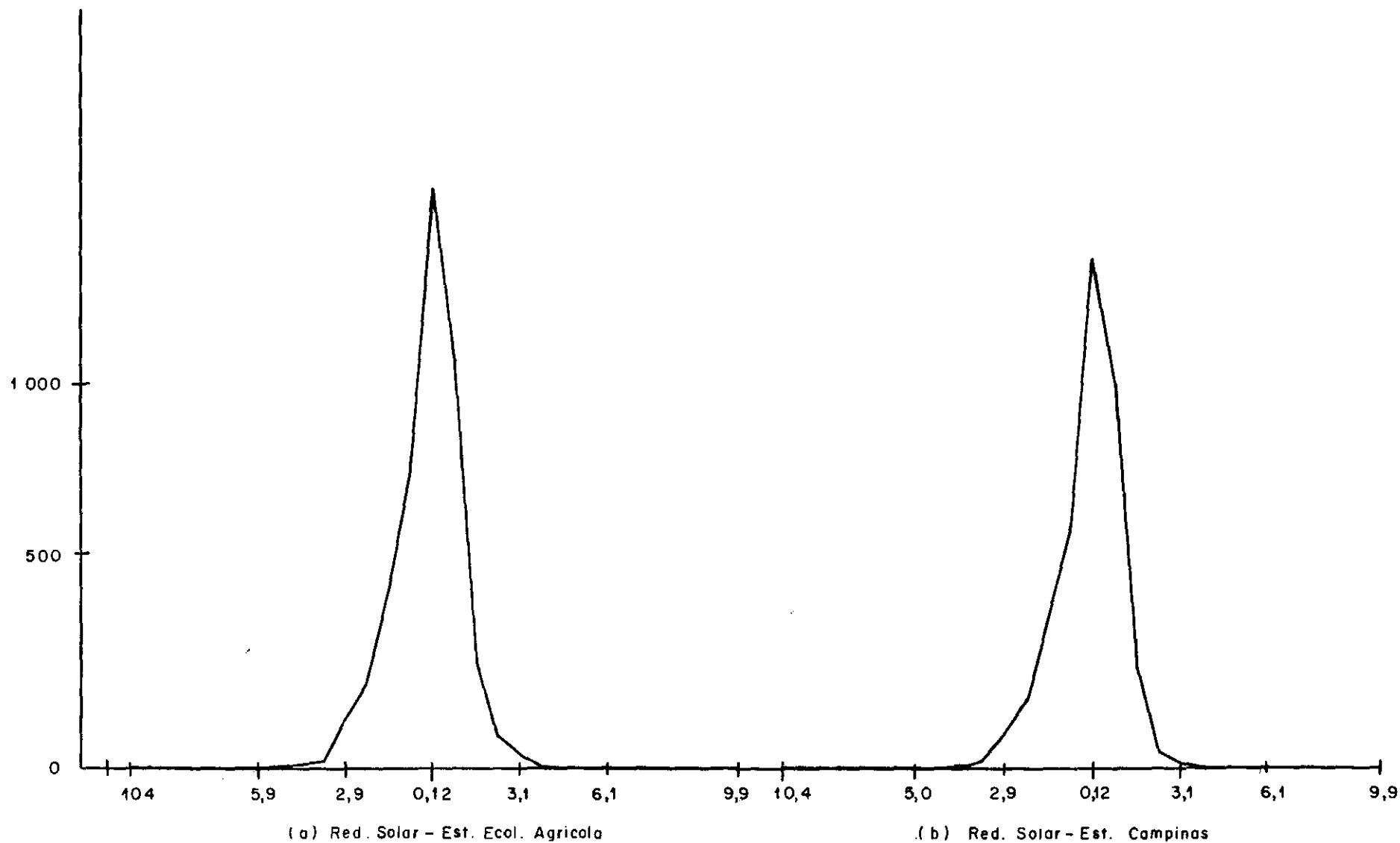
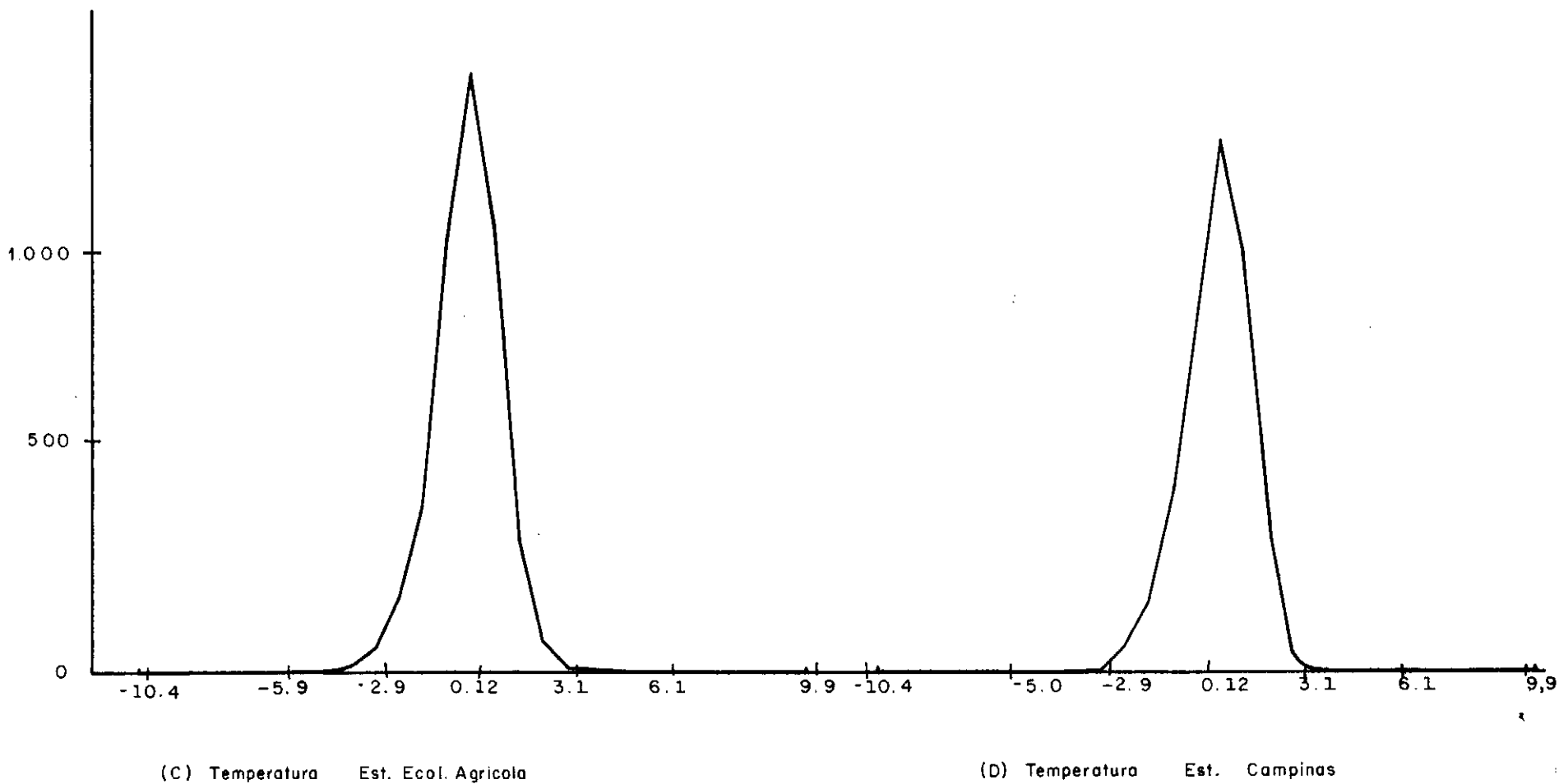


FIG. IV.22 - DISTRIBUIÇÃO DE FREQUÊNCIAS



V. CONCLUSÕES E RECOMENDAÇÕES

Os resultados obtidos permite as seguintes conclusões:

- Dentro da pouca disponibilidade de dados e das limitações apresentadas pelos testes estatísticos auxiliares na análise da estrutura das séries geofísicas temporais, considera-se os modelos selecionados como representativo das séries de radiação solar global e temperatura média do ar para as localidades de Campinas, São Paulo e Km 47, Rio de Janeiro.
- Os resultados dos testes de constancia de parâmetros devem ser interpretados cuidadosamente tendo em vista que as premissas em que são fundamentados conflitam com o caráter das séries geofísicas e nem sempre se tem uma justificativa física para a não aceitação da hipótese.
- Os métodos existentes para inferencia de harmônicos significantes, principalmente na variancia, se mostraram inadequados para as séries diárias de radiação solar global e temperatura média do ar analisadas, ficando a seleção sendo subjetiva.

- A dependencia na componente estocástica das séries de radiação solar global e temperatura média do ar podem ser modeladas por processos autoregressivos lineares, entretanto, não se pode garantir, *a priori*, que estes sejam os mais adequados.
- Na seleção da ordem dos modelos autoregressivos, o teste de Yevjevich não mostrou precisão suficiente. Neste caso, o teste de Siddiqui, mesmo apresentando limites muito estreitos, parece ser mais conclusivo.
- Não se conseguiu ajustar uma função de distribuição de probabilidade as distribuições de frequência da componente independente da radiação solar global e temperatura média do ar diárias. A distribuição de frequência acumulada pode representar estas distribuições na carencia de uma expressão funcional.

As seguintes recomendações podem ser indicadas em função da experiência adquirida:

- Realização de um trabalho prévio, exaustivo, da qualidade dos dados afim de evitar dúvidas quando na averiguação da presença ou ausencia de tendencias e/ou saltos.
- Desenvolvimento de critérios mais adequados para preenchimento das falhas.

- Aplicação da transformada rápida de Fourier (FFT) à série observada na tentativa de detectar mais convenientemente os harmônicos significantes no processo, e/ou auxiliar na identificação dos harmônicos significantes nos parâmetros estatísticos das séries, e talvez detectar outros períodos além do período anual.
- Desenvolvimento de critérios mais adequados para seleção da ordem dos modelos autoregressivos lineares de séries diárias de radiação solar global e temperatura média do ar.
- Utilização de modelos autoregressivos lineares bem selecionados para preenchimento de falhas existentes nos dados observados.
- Utilização de modelos autoregressivos não lineares para a componente estocástica dependente das séries de radiação solar global e temperatura média do ar na tentativa de melhores resultados.
- Utilização de horas de brilho solar em futuras tentativas de estimação da radiação solar global, empregando a análise da estrutura de séries meteorológicas descrita neste trabalho.

BIBLIOGRAFIA

- ANDERSON, R.L., "Distribution of the Serial Correlation Coefficient", Annals of Mathematical Statistics, Vol. 13, 1941.
- ANDERSON, T.W., "The Statistical Analysis of Time Series", J. Willey & Sons, 1971.
- BARTLETT, M.S., "Periodogram Analysis and Continuos Spectra" , Biometrika (37), 1950.
- BOX, G.E.P. e JENKINS, G.M., "Time Series Analysis - Forecasting and Contrl", Holden-Day, 1970.
- BRYVIN, M.C., "Heavy-Tailed Distributions, Properties and Tests", Technometrics, Vol.16, No.1, pp.61-67, 1974.
- CHAKRAVANTI, I.M., LAHA, R.G. e ROY, R., "Handbook of Methods of Applied Statistics, 2 Vol., John Wiley & Sons, Inc., 1,560 pp., 1967.
- COSTA, J.J.S., "Probabilidades e Processos Estocásticos", Universidade de São Paulo, Editora Vozes Ltda., 1971
- CLUIS, D., "Relation entre la température de l'eau d'une rivière et la temperature de l'air ambiant", Rap. Tech., INRS-EAU, Québec, 3, 11p., 1972.
- EID, N.J., "Probabilidades de duração de secas empregando a teoria de runs", Tese de Mestrado, COPPE/UFRJ, 1976.

- FISHER, R.A., "Tests of Significance in Harmonic Analysis",
 Proceedings of the Royal Society, Series A, Vol. 125, pp.
 55-59, 1929.
- FURMAN, R.Wm., "Time Series Analysis of Daily Maximum Temperature on a Mountain Slope", Journal of Applied Meteorology,
 Vol. 17, janeiro de 1978.
- GRIFFTH, H.L., PANOFISKY, H.A. e VAN DER HOVER, J., "Power Spectrum Analysis over large ranges of frequency, J.Meteorology, 13(3):279-282, 1956.
- GRENANDER, U. e ROSENBLATT, M., "Statistical Analysis of Stationary times series", J.Willey & Sons, 1957.
- HANNAM, E.J., "Multiple Time Series", J.Willey & Sons, 1970.
- JULIAN, P.R., "Variance Spectrum Analysis", Water Resour. Res., 3(3):831-845, 1967.
- KELMAN, J., "Stochastic Modeling of Intermittent Daily Hydrologic Series", Ph.D Dissertation, Colorado State University, Fort Collins Colorado, 1975.
- KOTHANDARAMAN, V., "Analysis of Water Temperature Variations in large rivers, Proc. Am. Soc. Civ. Eng., J.Sanit. Eng. Div. 97(SAI): 19-31, 1971.
- KOTTEGODA, N.T., "stochastic Five Day Streams Flow Model", Journal of the Hydraulics Division, Proc. Am.Soc. Civil Eng., Hy 9, 1972.

- LATTARI, L.A. e GUERRERO, P., "Modelagem Hidrológica para Estudos de Operação a longo prazo de sistemas hidrelétricos", Revista Brasileira de Energia Elétrica, nº 30, 1975.
- LLAMAS, J. e SIDDIQUI, M.M., "Runs of Precipitations Series", Colorado State University, Hydrology Paper No. 33, Fort Collins, Colorado, 1969.
- LLAMAS, J. e TRIBOULET, J.P., "La temperature de l'air et le cycle hydrologique: Structure Statistique", Centre de Recherches sur l'eau, Université Laval, Vol.1, No. 3, Canada, 1976.
- MACHADO, D.O., "Modelos Estocásticos para Análise e Simulação de Sequências Hidrológicas", Tese de Mestrado, COPPE/UFRJ, 1970.
- MARKOVIC, R.D., "Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff", Colorado State University, Hydrology Paper No. 8, Fort Collins, Colorado, 1965.
- MATALAS, N.C., "Mathematical Assessment of Synthetic Hydrology", Water Resources Research, Vol.3, No.4, pp.937-945, 1967.
- MITCHELL, J.M. et al., "Climatic Change", Tech. Note, OMM, Geneve, 79, WMO No. 195, TP. 100, 79 pp., 1966.
- NEVILLE, A.M. e KENNEDY, J.B., "Basic Statistical Methods", International Textbook Company, Scranton, Pennsylvania, 1964.

- PANOSFKY, H.A., "Meteorological Application of Power-Spectrum Analysis", Bull. Amer. Meteor. Soc., 36(4):163-166, 1955.
- PALTRIDGE, e PLATT, "Radiative Processes in Meteorology and Climatology", Ed. Elsevier, 1976.
- QUENOUILLE, M.H., "A large sample test for the godness of fit of autoregressive schemes", Journal Royal Statistical Society, V.110, 1949.
- QUIMPO, R.G., "Stochastic Model of Daily River Flow Sequences", Colorado State University, Hydrology Paper, no. 18, Fort Collins, Colorado, 1967.
- QUIMPO, R.G. e YEVJEVICH, V., "Stochastic description of daily river flows", Proceedings International Hydrology Symposium, Fort Collins, Colorado, Vol. I, p.290-297, 1967.
- RICHARDSON, C.W., "A model of stochastic structure of daily precipitation over air area", Ph.D Dissertation, 1976.
- RODRIGUEZ-ITURBE, I. et al., "Adequacy of Markovian Models with cyclic components for stochastic streamflow simulations", Water Resources Research (7) S., 1971.
- ROESNER, L.A. e YEVJEVICH, V., "Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff, Colorado State University Hydrology Paper, No. 15, Fort Collins, Colorado, 1966.

- SALES, P.R.H., "Um modelo determinístico-estocástico para previsão de séries hidrológicas temporais", Tese de Mestrado, COPPE/UFRJ, 1977.
- CHUSTER, A., "On the Investigation of Hidden Periodicities with Applications to a supposed 26 day period of a meteorological phenomena", Terr. Mag. Vol. (3), no. 1, (J. of Geo. Research).
- SERRA, A., "A Mean Temperature Formula", Serviço de Meteorologia, Brasil, 1963.
- SIDDIQUI, M.M., "Trabalho não publicado conforme citado em Yevjevich 1972b".
- SONG, C.C.S., PABST, A.F. e BOWER, C.E., "Stochastic Analysis of Air and Water Temperature", Proc. Am. Soc. Civ. Eng., J. Environ. Div., EEC: 785-800, 1973.
- STURGES, H.A., "The Chalice of a class interval", Journal of America Statistical Association, Vol. 21, p.65-66, 1926.
- TAO, Pen-Chih e outros, "Distributions of Hydrologic Independent Stochastic Components", Colorado State University, Hydrology Papers, No. 82, Fort Collins, Colorado, 1976.
- TUCKEY, J.W., "The sampling theory of power spectrum estimates", Symposium on Applications of Autocorrelation Analysis to Physical Problems, U.S. Office of Naval Research (Navexos P-735), 1950.

- TUCKEY, J.W., "Discussion Emphasizing the Connection Between Analysis of variance and spectral analysis", *Technometrics* (3) 2., 1961.
- VELZ, C.J. e GANNON, J.J., "Forecasting Heat Loss in Ponds and Streams", *J.WPCF*, 32(4):392-417, 1960.
- WALKER, G.T., "On Periodicity", *Royal Meteorological Society Journal*, V.SI, pp.337-346, 1925.
- WISER, E.H., "Persistence Models in Precipitation Analysis", Ph.D Dissertation, North Carolina State University at Raleigh, 1964.
- YEVJEVICH, V., "Fluctuations of Wet and Dry Years", Part I, Research Data Assembly and Mathematical Models, Colorado State University Hydrologic Papers, No. 1, Fort Collins, Colorado, 1963.
- YEVJEVICH, V., "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation", Colorado State University, Hydrology Papers, No. 4, Fort Collins, Colorado, 1964
- YEVJEVICH, V., "Stochasticity in Geophysical and Hydrological Time Series", *Nordic Symposium on Stochastic Hydrology*, Uppsala, Sweden, September, 1971.
- YEVJEVICH, V., "Structural Analysis of Hydrologic Time Series", Hydrology Paper, No. 56, Colorado State University, 1972a.
- YEVJEVICH, V., "Stochastic Processes in Hydrology", *Water Resources Publications*, Fort Collins, Colorado, 1972b.

ANEXO IESTIMADORES DAS ESTATÍSTICAS DAS AMOSTRAS

MÉDIA:

$$m = \frac{1}{N} \sum_{p=1}^N X_p$$

VARIANCIA:

$$s^2 = \frac{1}{N} \sum_{p=1}^N (X_p - m)^2 \quad \text{e quando } N < 30, \quad s^2 = \frac{N}{N-1} s^2$$

DESVIO PADRÃO:

$$s = \left(\frac{1}{N} \sum_{p=1}^N (X_p - m)^2 \right)^{1/2} \quad \text{e quando } N < 30, \quad s = s \frac{N}{N-1}$$

COEFICIENTE DE ASSIMETRIA:

$$g_1 = \frac{1}{N} \sum_{p=1}^N (X_p - m)^3 / (s^2)^{3/2} \quad \text{e quando } N < 30, \quad g_1 = g_1 \frac{N^2}{(N-1)(N-2)}$$

COEFICIENTE DE CURTOSE:

$$g_2 = \frac{1}{N} \sum_{p=1}^N (X_p - m)^4 / (s^2)^2 \quad \text{e quando } N < 30, \quad g_2 = \frac{N^3}{(N-1)(N-2)(N-3)}$$